University of Warsaw
Faculty of Physics

Julia Romanowska

Comparing physical properties of aminoglycoside antibiotics' binding sites in RNA and proteins

PhD dissertation

Supervisor:
Joanna Trylska, PhD, Assoc. Prof.
Centre of New Technologies,
University of Warsaw

Warsaw
June 2012

Abstract

Aminoglycoside antibiotics have been in use for more than 60 years, helping combat severe bacterial infections. Due to this long time of usage, more and more bacteria become resistant to one or several drugs from this group. This spread of resistant species is alarming and additionally, there is little knowledge about the mechanisms of bacterial resistance.

In order to broaden our understanding of how bacteria combat aminoglycosides, we performed computer simulations of various molecules that bind aminoglycosides in a bacterial cell: *(i)* the primary binding site, called the A-site and located in ribosomal RNA, wild type and with mutations that decrease the aminoglycoside binding affinity; and *(ii)* the aminoglycoside modifying enzymes (AMEs), which are produced by bacteria to inactivate these drugs. The mutations of the RNA A-site were chosen based on previous experimental studies on whole bacteria. These studies showed that even single base substitutions were sufficient to make bacteria resistant, but did not explain how this resistance was gained on an atomic level. There are many AMEs and they vary a lot among themselves, yet they all have a narrow specificity towards aminoglycosides, which are quite homogeneous group. The two main questions we have posed in our research are: *(i)* what are the physical grounds of bacteria becoming less susceptible to aminoglycosides due to RNA A-site mutations; and *(ii)* how different AMEs attract aminoglycosides and interact with them?

We performed all-atom molecular dynamics (MD) simulations of the A-site model with selected mutations and of AME representatives. In addition, the complexes of these biomolecules with aminoglycosides were simulated. For comparison, we also performed simulations of the wild type A-site model and of the aminoglycosides in water. We used various biophysical methods to analyze these simulations and to study: internal dynamics of the biomolecules; electrostatic potential, shape, and volume of the binding pockets; types of interactions with aminoglycosides; and changes in conformations of aminoglycosides. In addition, we developed and implemented an algorithm that helps describe molecular motions.

We found that different A-site mutations affect different features of the RNA binding site. Some of them changed the mobility of the nucleic bases, and therefore the shape of the A-site was altered. Other mutations changed the electrostatic potential inside the binding site, thus making it almost unrecognizable to aminoglycosides. The study of AMEs showed that apart from their structural and sequence-related diversity, they differ in the internal movement patterns. However, these enzymes interact with aminoglycosides very similarly, using mainly electrostatic interactions. Interestingly, we noticed that these interactions were copied from the RNA:aminoglycoside complex. Our findings were in agreement with experimental studies and also helped to explain some of their outcomes. The results presented in this dissertation may help design new antibiotics that would overcome the bacterial resistance.

Dissertation title and summary in Polish

Analiza porównawcza własności fizycznych miejsc wiązania antybiotyków aminoglikozydowych w RNA i białkach

Od ponad 60 lat antybiotyki aminoglikozydowe są z powodzeniem stosowane w szpitalach przeciwko ciężkim infekcjom bakteryjnym. Jednak pojawianie się coraz większej liczby przypadków bakterii opornych na stosowane aminoglikozydy sprawia, że badania mechanizmów oporności u bakterii stają kluczowe w dalszej skutecznej walce z infekcjami tego typu.

Przeprowadziłam komputerowe symulacje biomolekuł, które oddziałują z antybiotykami aminoglikozydowymi we wnętrzu komórek bakteryjnych. Badanymi obiektami są: *(i)* główne miejsce wiązania aminoglikozydów, zwane miejscem A, w rybosomalnym RNA; natywne oraz z mutacjami powodującymi wzrost oporności u bakterii; a także *(ii)* enzymy modyfikujące aminoglikozydy (ang. *aminoglycoside modifying enzymes, AME*), produkowane przez bakterie w celu chemicznej dezaktywacji tych leków. Motywacją do badań nad zmutowanym miejscem A był brak informacji o zmianach jakie zachodzą w fizycznych własnościach miejsca A po różnych zamianach nukleotydów. Wiadomo jakie mutacje prowadzą do oporności oraz że nawet pojedyncze zamiany nukleotydu mogą mieć bardzo wymierne skutki, ale nie wyjaśniono jakie są tego podstawy. Natomiast, w przypadku AME, celem prowadzenia symulacji było wyjaśnienie w jaki sposób ta grupa białek jest w stanie być jednocześnie bardzo zróżnicowana i wysoce specyficzna względem aminoglikozydów.

Przeprowadziłam symulacje dynamiki molekularnej (MD) modelu miejsca A z wybranymi mutacjami oraz reprezentatywnych enzymów z trzech największych rodzin AME. Aby uzyskać opis oddziaływań między tymi miejscami wiążącymi a aminoglikozydami, przeprowadziłam również symulacje MD tych biomolekuł w kompleksach z wybranymi antybiotykami. W celu analizy symulacji użyłam metodologii z zakresu biofizyki teoretycznej. Badałam wiele własności fizykochemicznych wybranych biomolekuł i ich kompleksów, m.in.: dynamikę wewnętrzną, własności elektrostatyczne, kształt i objętość miejsc wiązania aminoglikozydów, a także rodzaje oddziaływań z aminoglikozydami. Ponadto, stworzyłam nową metodę analizy zmian konformacyjnych w molekułach, która dokonuje podziału biomolekuł na tzw. dynamiczne domeny, na podstawie danych pochodzących z symulacji lub eksperymentów.

Z analizy symulacji rybosomalnego miejsca A wynika, że mutacje różnych zasad wpływają na różne własności fizyczne tego fragmentu RNA. W zależności od położenia mutowanej zasady, zmieniał się rozkład ładunków cząstkowych w miejscu wiążącym lub kształt tego miejsca. Mutacje wpływały również na dynamikę ruchów wewnętrznych miejsca A. Analiza symulacji cząsteczek AME wskazała, że oprócz różnorodności struktur trzeciorzędowych i sekwencji, występuje w tej grupie również różnorodność w ruchach wewnętrznych. Pomimo tych różnic, wszystkie enzymy oddziaływały z aminoglikozydami w bardzo podobny sposób, głównie elektrostatycznie. Ponadto, te oddziaływania wydają się być kopiowane z kompleksów, jakie aminoglikozydy tworzą z miejscem A. Rezultaty moich badań są zgodne z poprzednimi doniesieniami eksperymentalnymi, a także pomagają wyjaśnić niektóre z nich. Wyniki opisane w tej pracy mogą być podstawą do zaprojektowania zmodyfikowanych aminoglikozydów, które mogłyby być aktywne nawet wobec opornych bakterii.

Acknowledgements:

I have met many wonderful people during this scientific journey towards a PhD degree and I am grateful for all the nice words I heard and the atmosphere these people created. I would definitely not have done it without my family, my dearest Lars and my faithful friends. It would not have been that fun without my colleagues from the labs: Bionano, McCammon's group and Reuter's group. But most of all, it would not be possible at all without the encouragement, enormous amount of help and guidance, and sometimes very necessary but always constructive criticism from my super supervisor, Joanna.

Thank you.

Funding:

The research was supported by ICM, University of Warsaw (G31-4), Polish Ministry of Science and Higher Education (N N301 245236 and N N301 033339), Foundation for Polish Science (Focus program and Team project (TEAM/2009-3/8) co-financed by European Regional Development Fund operated within Innovative Economy Operational Programme), the "HPC Infrastructure for Grand Challenges of Science and Engineering" Project, co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme. The work at University of Bergen was co-financed by technical assistance funds of the EEA Financial Mechanism and the Norwegian Financial Mechanism within the framework of the Scholarship and Training Fund.

List of publications

1. Romanowska, J., Reuter, N. & Trylska, J. (2012) Comparing aminoglycoside binding sites in their bacterial modifying enzymes and ribosomal RNA, *Proteins*, preparing revised manuscript (after Peer Review).

2. Romanowska, J., Nowiński, K. S. & Trylska, J. (2012) Determining geometrically stable domains in molecular conformation sets, *J. Chem. Theory Comput.*, preparing revised manuscript (after Peer Review).

3. Romanowska, J., McCammon, J. A. & Trylska, J. (2011) Understanding the Origins of Bacterial Resistance to Aminoglycosides through Molecular Dynamics Mutational Study of the Ribosomal A-site. *PLoS Comp. Biol.*, **7**, e1002099.

4. Romanowska, J., Ekonomiuk, D. & Trylska, J. (2011) Computational studies of RNA dynamics and RNA-ligand interactions. Tor, Y. & Wanunu, M. (eds.), *Methods for Studying Nucleic Acid Drug Interactions*, chap. 15, Taylor & Francis.

5. Romanowska, J., Setny, P. & Trylska, J. (2008) Molecular dynamics study of the ribosomal A-site. *J. Phys. Chem. B.*, **112**, 15227–15243.

Selected talks and poster presentations

1. *Mar. 26–29, 2007; Heidelberg, Germany*; **Biological Diffusion and Brownian Dynamics Brainstorm**; poster presentation: "Dynamics of aminoglycosidic binding site in the small ribosomal subunit";

2. *Mar. 28–Apr. 3, 2008; Steamboat Springs, Colorado, USA*; **Computer-Aided Drug Design**; poster presentation: "Molecular dynamics study of the ribosomal A-site";

3. *Apr. 24–27, 2008; Warsaw, Poland*; **Bioinformatics 2008**; poster presentation: "Molecular dynamics study of the ribosomal A-site";

4. *Oct. 3–5, 2008; Jadwisin, Poland*; **First Convention of the Polish Bioinformatics Society**; oral presentation: "Molecular dynamics study of the ribosomal A-site";

5. *Feb. 21–26, 2009; Zakopane, Poland*; 36th Winter School of the Faculty of Biochemistry, Biophysics and Biotechnology of the Jagiellonian University **Molecule interactions in health and disease**; oral presentation: "Molecular dynamics study of the ribosomal A-site";

6. *Sept. 30, 2009; San Diego, CA, USA*; **prof. McCammon's group meeting** (UCSD, Dept. of Pharmacology); oral presentation: "Characteristics of chosen aminoglycosidic antibiotics binding sites";

7. *May 14–16, 2010; University of Warsaw, Poland*; **Advanced Bioinformatics Tools**; poster presentation: "A variety of aminoglycoside complexes — in search of common binding site features";

8. *Jun. 13–16, 2010, Cetraro, Italy*; **International Society of Quantum Biology and Pharmacology, 2010 President's Meeting**; poster and oral presentation: "Insights into aminoglycosidic antibiotics' recognition mechanisms";

9. *Sept. 28, 2010; Bergen, Norway*; **dr. Nathalie Reuter's group meeting** at University of Bergen; oral presentation: "Comparing physicochemical features of biomolecules interacting with aminoglycosidic antibiotics by means of computational modeling";

10. *Mar. 29, 2011; Anaheim, CA, USA*; **241st American Chemical Society National Meeting**; oral presentation: "Insights into aminoglycoside antibiotics recognition mechanisms";

11. *May 4, 2011; University of Warsaw, Poland*; **Research Seminar "Mathematical Methods in Biology"** at Faculty of Mathematics, Informatics and Mechanics; oral presentation: "Physicochemical features of aminoglycoside modifying enzymes";

12. *May 22–26, 2011; Stockholm, Sweden*; **IXth European Symposium of The Protein Society**; poster presentation: "Understanding aminoglycoside recognition by aminoglycoside modifying enzymes and ribosomal RNA"; awarded *Protein Science Young Investigator Travel Grant*;

13. *Sept. 13–16, 2011; Zurich, Switzerland*; CECAM workshop: **Dynamics of Protein-Nucleic Acid Interactions: Integrating Simulations with Experiments**; poster presentation: "Understanding aminoglycoside recognition by aminoglycoside modifying enzymes and ribosomal RNA";

14. *Oct. 4–5, 2011; International Institute of Molecular and Cell Biology, Warsaw, Poland*; **The Modern Techniques for Drug Design Purposes**; poster presentation: "Understanding aminoglycoside recognition by aminoglycoside modifying enzymes and ribosomal RNA";

15. *Nov. 16–19, 2011; Nałęcz Institute of Biocybernetics and Biomedical Engineering, Warsaw, Poland*; **Multi-Pole Approach to Structural Biology**; poster presentation: "Understanding aminoglycoside recognition by aminoglycoside modifying enzymes and ribosomal RNA".

Contents

List of Figures

List of Tables

Abbreviations, in alphabetical order

2-DOS = 2-deoxystreptamine;
A-site = aminoacyl-tRNA binding site in the ribosome;
AAC = aminoglycoside acetyltransferase;
ACO = acetyl-coenzyme A;
ADP, AMPCPP, ATP = adenosine diphosphate, adenosine 5′-methylenediphosphate, adenosine triphosphate;
AG = aminoglycoside;
AME = aminoglycoside modifying enzyme;
AMSM = atomic movement similarity matrix;
ANT = aminoglycoside nucleotidyltransferase;
AO = atomic orbitals;
APH = aminoglycoside phosphotransferase;
APO state = here, non-active binary complex enzyme:cofactor;
CG = conjugate gradients (minimization method);
DCCM = dynamic cross-correlation matrix;
DNA = deoxyribonucleic acid;
ESP = electrostatic potential;
GB = generalized Born;
GDP, GTP = guanosine diphosphate, guanosine triphosphate;
GTO = Gaussian-type orbital;
HOLO state = here, active ternary complex enzyme:cofactor:aminoglycoside;
KAN = kanamycin A;
LCAO = linear combination of atomic orbitals;
MD = molecular dynamics;
MIC = minimal inhibitory concentration;
MMGBSA, MMPBSA = molecular mechanics general Born surface area, molecular mechanics Poisson-Boltzmann surface area;
MO = molecular orbital;
NMA = normal mode analysis;
NMR = nuclear magnetic resonance;
PAR = paromomycin;
PB = Poisson-Boltzmann (equation);
PCA = principal component analysis;
PDB = Protein Data Bank;
PES = potential energy surface;
PME = particle mesh Ewald;
QM = quantum mechanics;
RESP = restrained electrostatic potential method;
RMSD = root-mean-square deviation;
RMSF = root-mean-square fluctuation;
RNA, mRNA, rRNA, tRNA = ribonucleic acid, messenger RNA, ribosomal RNA, transfer RNA;
S = svedberg, a measure of sedimentation time of a molecule: $1\ \mathrm{S} = 10^{-13}$ seconds $= 100$ fs;
SASA = solvent accessible surface area;
SD = steepest descent (minimization method);
WT = wild-type.

Nucleic acid bases:
A, Ade = adenine;
C, Cyt = cytosine;
G, Gua = guanine;
U, Ura = uracil.

Amino acid names used in the text:
Arg, R = arginine;
ASH = protonated aspartic acid;
Asp, D = aspartic acid;
Asn, N = asparagine;
Gln, Q = glutamine;
Glu, E = glutamic acid;
Gly, G = glycine;
His, H = histidine;
Ile, I = isoleucine;
Lys, K = lysine;
Phe, F = phenylalanine;
Ser, S = serine;
Thr, T = threonine;
Trp, W = tryptophan;
Tyr, Y = tyrosine.

Chapter 1

Introduction

Bacterial cells are typically only several micrometers in size and visible only under a microscope. Yet there are millions of these cells in every inch of all that surrounds us. Some bacteria can cause a number of infectious diseases that can be fatal, while others are necessary for human health. Since the 19th century discovery of the connection between certain diseases and bacteria, there has been constant fight between bacterial survival and human attempts to eradicate pathogenic bacterial species. The first antibiotic, penicillin, was discovered by Alexander Fleming and was initially highly successful in controlling bacteria. Evolution, however, enabled bacterial species to resist the antibiotics.[1] Their short proliferation time, sometimes as short as minutes,[2] and their ability to mutate and incorporate external DNA, makes their resistance just a matter of time. Recent discoveries point to a possibility that some tools used by bacteria to subvert antibiotic action were available almost from the beginning of life.[3]

Early antibiotic studies can be called *discoveries* — among numerous substances a researcher could only hope to find the one that would be lethal to bacteria while having minimal effects towards humans. Today, *drug design* is used to design inhibitory molecules that are based on knowledge of the biochemical processes in a bacterial cell. Drug design requires many in-depth studies of the basic mechanisms underlying the modes of action and resistance towards specific antibiotics in order to explain and clarify the interactions of the drug inside a bacterial cell and to propose development strategies.

While the final stages of drug development cannot be completed without experiments on living organisms, modern computational techniques can substantially reduce the time and cost of the initial stages of drug development. Using the modern tools of theoretical biophysics, the interactions between antibiotic molecules and their receptors in bacterial cells can be reliably investigated with computational modeling.[4,5] Some of these methods enable detailed studies of static structures, but most importantly, the dynamic features of biomolecules can be directly observed and measured. Theoretical approaches are now commonly used in research related to drug design.[6]

Notably, drug development is not only about fighting new diseases. There is a constant need for better antibacterial agents due to the bacterial resistance that develops against widely used drugs. This is especially important for antibiotics used only in hospitals against the multidrug resistant bacteria species. One example is the treatment of multi-drug resistant tuberculosis with aminoglycoside antibiotics (AGs),[7] which are still successfully used but the resistance to these antibiotics is spreading. According to the World Health Organization (WHO), in some regions more than 18% of new tuberculosis cases are multi-drug resistant.[8] Therefore, there is a pressing need to understand how bacteria fight these antibiotics, so that better drugs can be designed.

1.1 Aminoglycoside antibiotics

The history of aminoglycosides began with the discovery of streptomycin in the 1940s in the laboratory of Dr Selman Abraham Waksman.[9] Since then, aminoglycosides have been successfully used against severe bacterial infections. This class of antibiotics is active against many species of Gram-positive and Gram-negative bacteria.[10] Streptomycin was the first efficient weapon against deadly tuberculosis, which was an important breakthrough and brought its discoverer the Nobel prize in 1952.

From the chemical point of view, aminoglycosides are sugar (2-deoxystreptamine, 2-DOS) derivatives and almost all have a common core, neamine (Figure 1.1). These drugs are positively charged in physiological pH (ca. 7.4), due to the presence of many NH_2 groups, which become then NH_3^+.[11,12] There are two subclasses of 2-DOS aminoglycosides: 4,5-disubstituted and 4,6-disubstituted, which are also called neomycin-like and kanamycin-like, respectively. While these groups share many features, they also have some differences e.g., in resistance profiles.

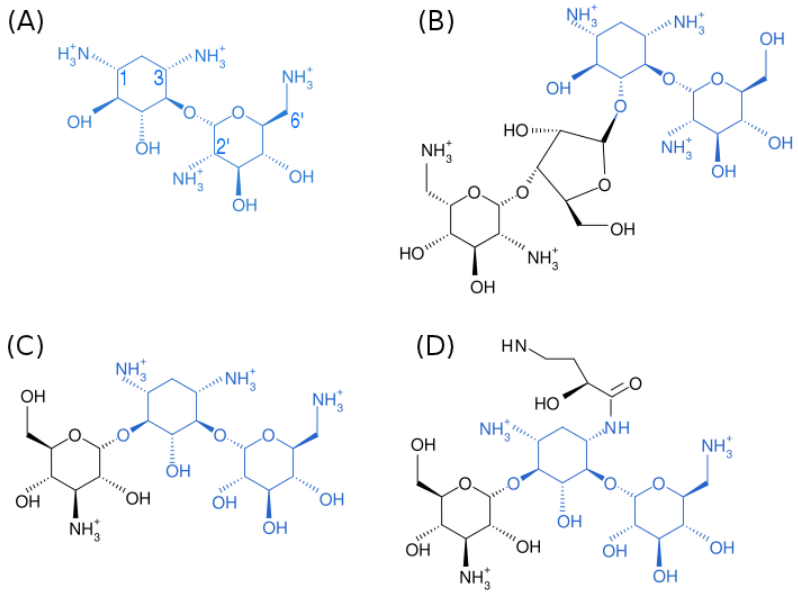


Figure 1.1: **Chemical structures of exemplary aminoglycosides**: (A) neamine, the common core of most aminoglycosides; the positions of the amine groups are numbered; (B) paromomycin (4,5-disubstituted 2-DOS); (C) kanamycin and (D) amikacin (both are 4,6-disubstituted 2-DOS).

1.1.1 Mode of action

The primary binding site of aminoglycoside antibiotics is located in the bacterial ribosome. Ribosomes are macromolecular complexes of proteins and RNA[13] that provide a framework for translation process in every living cell (Figure 1.2A). Ribosomes consist of two parts: the small (called 30S in prokaryota, 40S in eukaryota) and large (50S in prokaryota, 60S in eukaryota) subunits, with the names based on subunit sedimentation coefficients. In prokaryotic organisms, the large subunit consists of two RNA chains (23S and 5S) and ca. 30 proteins, while the small subunit has ca. 20 proteins and only one RNA chain (16S). At the interface of these two subunits, there are three binding sites for tRNA molecules, which provide amino acids according to the mRNA template information. The majority of aminoglycosides target one of the tRNA binding sites, namely the part of the A-site that is in 16S

rRNA, i.e., in the small ribosomal subunit[14] (Figure 1.3).

The sequence of the 16S rRNA A-site (Figure 1.3B) is highly conserved among different species and displays high similarity even between bacterial and human sequences (ca. 80%).[15,16] This similarity is responsible for the wide spectrum of aminoglycosidic action, however, it also enables these antibiotics to bind to human ribosomes, which can result in toxicity.[17] The most conserved residues in the A-site sequence are two adenines: A1492 and A1493 (numbering as in *Escherichia coli* is used throughout the text). These residues form a bulge in this helical rRNA fragment, and in the native ribosome they are very mobile, alternating between the flipped-out and flipped-in states[18,19] (Figure 1.4). This mobility is needed for the correct recognition of a tRNA molecule that carries an amino acid matching the mRNA codon.[20–24] When A1492 and A1493 are in the flipped-out state, they interact with other parts of the 16S rRNA, and send a signal that the tRNA can be accepted by the ribosome. When an aminoglycoside is bound, the adenines are locked in the flipped-out state[11,25–27] (Figure 1.3C), and therefore the acceptance ratio is almost the same for any tRNA, without differentiation between the cognate, near- and non-cognate. Consequently, the resulting proteins are malfunctioning, which eventually leads to bacterial cell death (Figure 1.2B).

Apart from this diminished translation fidelity, some aminoglycosides block the correct assembly of the ribosomal subunits.[9] More recently, another aminoglycoside feature has drawn attention, wherein binding of some aminoglycosides, e.g., paromomycin, to the ribosome allows the translation to proceed past the AUG stop codon on mRNA. This function, termed a stop-codon read-through ability,[28,29] could be useful in treating genetic diseases caused by a premature stop codon mutation in humans, wherein ribosomes create truncated proteins that do not function properly. Several modified aminoglycosides have been proposed that could act in human cells[29] to restore the proteins to their native length.

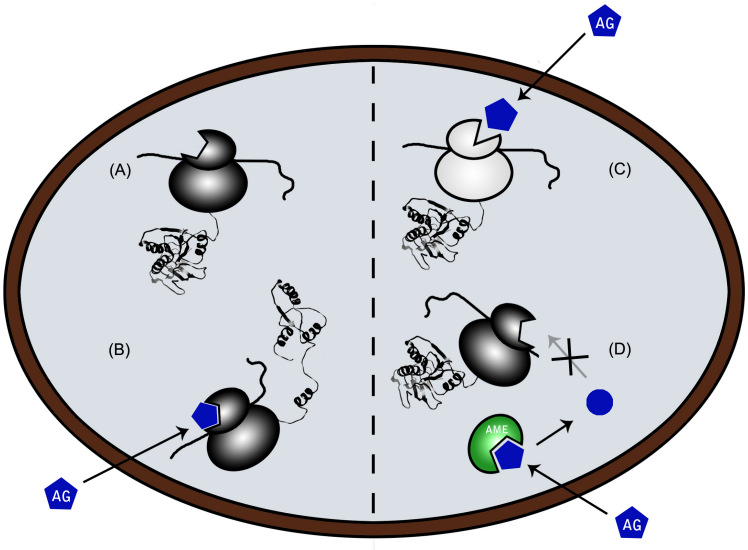


Figure 1.2: **Scheme of aminoglycoside action and bacterial resistance**. *(A)* Normal translation process; *(B)* binding of an aminoglycoside (AG) to the 30S subunit increases the translation error rate, thus producing malfunctioning proteins; *(C)* specific mutations within the aminoglycoside binding site result in diminished drug affinity; *(D)* production of aminoglycoside modifying enzymes (AME) by bacteria; the chemically altered antibiotics are not able to bind to the ribosomal binding site.

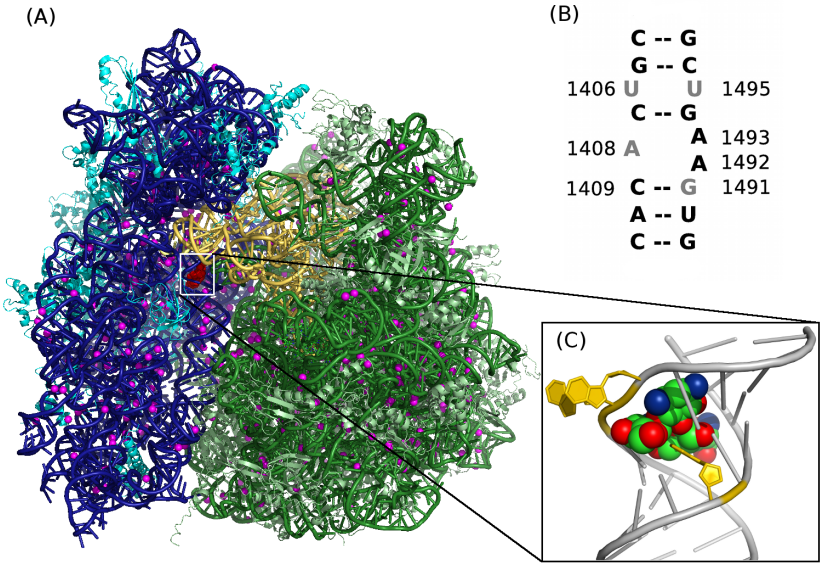


Figure 1.3: **Paromomycin binding site in the bacterial ribosome**: *(A)* cartoon representation of the entire ribosome during the translation process, with the small subunit in blue (16S rRNA) and cyan (proteins), and the large subunit in green (5S and 23S rRNA) and lime (proteins), the tRNA molecules are yellow and paromomycin is presented as red spheres; *(B)* the sequence of the aminoglycoside binding site in the *E. coli* 16S rRNA; *(C)* paromomycin (shown as spheres) bound to 16S rRNA, the bases marked yellow are A1492, A1493 (both in the flipped-out state) and A1408.

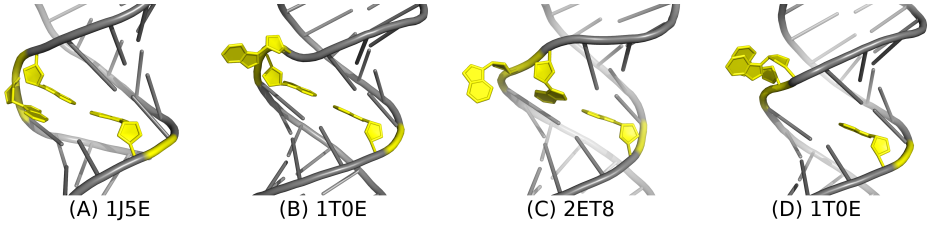


Figure 1.4: **Different conformational states of A1492 and A1493**, presented as a visualization of several X-ray structures of the A-site model or the entire 30S subunit. Each structure is identified by its PDB code (1J5E,[30] 1T0E,[26] 2ET8[31]).

1.1.2 Off-target binding sites

Inside the cell, aminoglycosides have a high net positive charge that causes a natural attraction to negatively charged molecules inside a cell, mainly nucleic acid chains. For example, it has been shown that aminoglycosides can bind and stabilize DNA triplexes.[32] However, it is mainly RNA that interacts with these drugs inside living cells[33] and only specific sequences can effectively bind these antibiotics.[34–36] These specific RNAs include ribozymes (i.e., RNA fragments that are capable of RNA cleavage), aptamers (also called RNA switches) or even HIV mRNA fragments.[37,38] This binding results in a specific function, with aminoglycosides affecting the hammerhead ribozyme cleavage[39] and inhibiting viral replication by blocking the necessary conformational changes of viral mRNAs.[37]

1.1.3 Bacterial resistance

Although aminoglycosides can bind to several different sites in various RNA chains, their practical use shows that the efficacy in binding to the primary ribosomal binding site is sufficient to combat bacterial infections. Even though the majority of patients currently

respond well to aminoglycoside treatment, more and more cases of resistant bacterial strains are being reported in clinics worldwide.[9] One measure of the resistance level is the minimal inhibitory concentration (MIC; Chapter 2 in Ref.[40]), which is the lowest concentration of a drug (in μg/ml) needed to completely stop bacterial growth, as determined after a minimum 16 h incubation. For susceptible bacteria, the aminoglycoside MIC is around 5 μg/ml but for resistant strains the MIC can increase by 1000-fold or more (e.g., Refs.[41,42]).

There are three main modes of the bacterial resistance to aminoglycoside antibiotics:[14,28] *(i)* decreasing the number of drug molecules inside the cell (via active and passive efflux); *(ii)* changing the primary binding site through mutation or methylation of A-site RNA bases (Figure 1.2C); and *(iii)* enzymatic chemical modification of the drugs (Figure 1.2D). The last two modes are the main focus of this thesis, and are introduced in the two following sections.

1.1.4 A-site modifications

Aminoglycosides target the very basic process in bacterial cells, thus the majority of the spontaneous mutations within the binding site are lethal for bacteria.[43,44] Moreover, there are usually several copies of genes encoding 16S rRNA[14] so that even when one RNA copy carries a resistance-causing sequence, the others remain susceptible. Experiments have shown that only specific mutations result in living resistant bacteria with even single-point mutations being sufficient to promote survival[41,42,45–47] (Table 1.1 presents exemplary mutations together with their MIC values).

Table 1.1: **Level of resistance caused by exemplary mutations in the bacterial 16S rRNA.** See Figure 1.3B for base numbering.

	MIC [μg/ml]			
mutation	**paromomycin**	**kanamycin A**	**amikacin**	**reference**
WT	1	1	1	[41]
A1408G	64	>1024	>1024	[41,48]
G1491A	32	2	1	[41,48]
G1491U	512	128	32	[41,48]
U1406C	64	32–64	4	[41,49]
U1495A	256–512	512	128–256	[41,49]
U1406C/U1495A	>1024	>1024	1024	[41,49]
U1406C/U1495G	>1024	>1024	128	[41,49]

The most important difference between bacterial and human A-site sequences is at base 1408, which is adenine (A) in prokaryota and guanine (G) in eukaryota.[50] Bacteria with an A1408G mutation are highly resistant (i.e., having a broad resistance profile) and their proliferation rate is almost the same as for the wild type.[51] There are many other mutations that give high levels of resistance, however, they also lead to diminished translation efficiency, which prolongs the time of bacterial growth. Moreover, not every mutation is equally efficient against every aminoglycoside — different distribution patterns of OH and NH_2 substituents in aminoglycosides correlate with different resistance profiles of bacteria (e.g., Refs.[41,42,52]). Nevertheless, the double mutation U1406C/U1495A gives a very broad resistance profile.[45] Interestingly, a similarly broad profile was found in bacteria possessing only one of these two substitutions, namely U1495A, while the other single mutation, U1406C, did not produce significant effects. Although some mutations result in high levels of resistance, they are not common among bacteria. Researchers have demonstrated that the most frequently occurring mutations are also the most profitable,[51] i.e., they decrease the efficiency of aminoglycosides significantly while not slowing down bacterial growth.

Nevertheless, only a few bacterial species can use this resistance method, namely these species that possess only one copy of the gene encoding the 16S rRNA chain. More recently, post-transcriptional methylation of this RNA fragment was shown to cause resistance in several Gram-negative pathogens.[53,54] This is an even smaller change than a single-point mutation but it still can produce high levels of resistance (MIC $\geqslant$ 256 μg/ml[53]), albeit against fewer aminoglycosides.

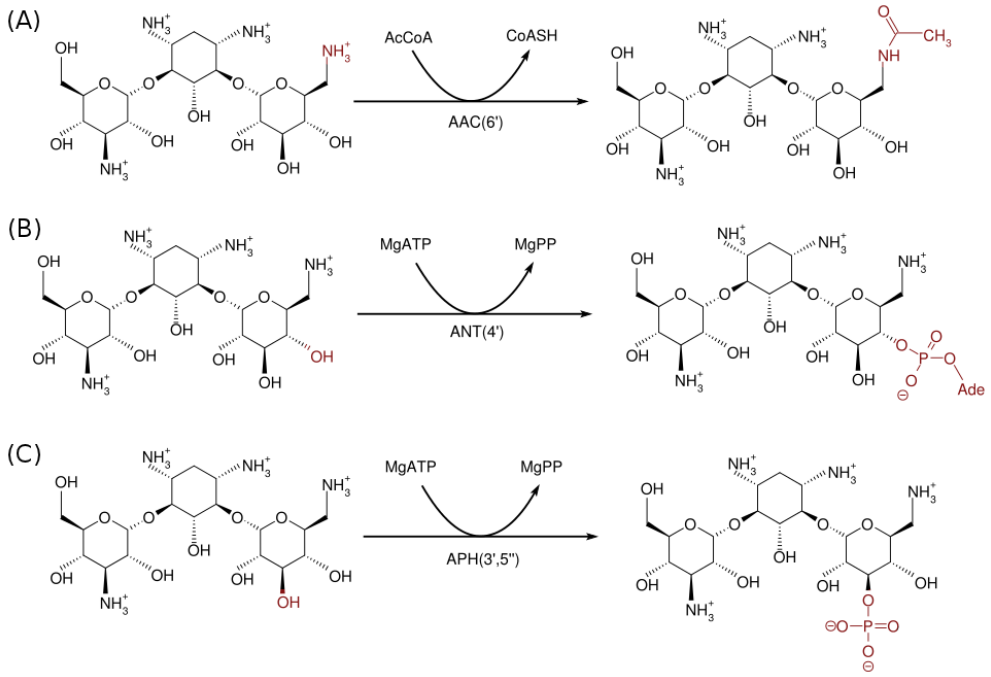


Figure 1.5: **Schemes of reactions catalyzed by representative AMEs** from different families: *(A)* acetyltransferase AAC(6′), *(B)* nucleotidyltransferase ANT(4′), and *(C)* phosphotransferase APH(3′,5″). The aminoglycoside being modified is kanamycin A and the affected substituents are marked in red.

1.1.5 Aminoglycoside modifying enzymes

The most commonly occurring method of counteracting aminoglycoside action is for bacteria to produce proteins that chemically modify these drugs, aminoglycoside modifying enzymes (AMEs).[14,55] Thus modified drugs cannot bind to the primary binding site in 16S rRNA.[56] There are a plethora of these enzymes[55] since the majority are chromosomally encoded and specific for one bacterial strain. In general, all AMEs can be divided into three families, depending on the type of reaction they catalyze: *(i)* (aminoglycoside) acetyltransferases (AACs), *(ii)* nucleotidyltransferases (ANTs), and *(iii)* phosphotransferases (APHs). These enzymes require a cofactor to work: nucleoside triphosphate (commonly ATP) for APHs and ANTs, and acetyl-coenzyme A (ACO) for AACs. Additionally, ANTs and APHs need divalent ions (most often Mg^{2+}) bound in the vicinity of the cofactor phosphate groups. These families are further divided into groups and types, depending on which aminoglycoside substituent they modify and from which bacterial strain they are derived. For example, APH(6)-Ia modifies the 6-OH group of ring II, and originates from *Streptomyces griseus* [Distler et al., 1987 in Ref.[55]]. Figure 1.5 depicts how one aminoglycoside, kanamycin A, can be modified by different AMEs.

Notably, these three AME families are not related in terms of their sequence, structure

or evolution. With regard to structure, the AAC family can be assigned to the GCN5-related N-acetyltransferase (GNAT) superfamily,[57] while the function of APH enzymes links them to the protein kinase family,[58] although the sequence similarity between these two groups is very low. ANTs are unrelated to any other known enzymatic group. Below, we present the current state of knowledge for three selected enzymes that represent the different families: AAC(6′)-Ib, ANT(4′)-I and APH(3′,5″)-IIIa.

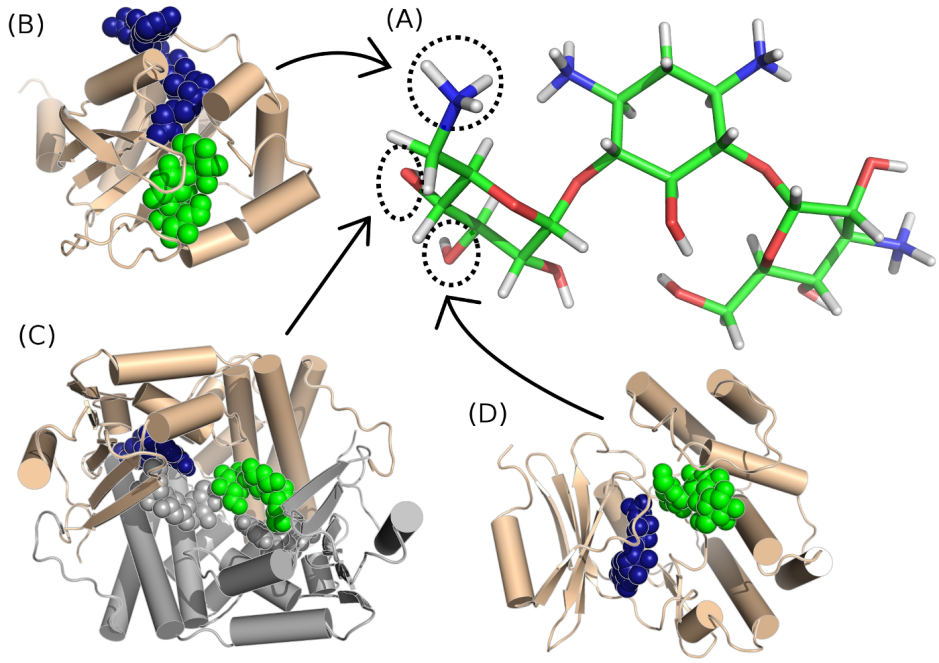


Figure 1.6: **Visualization of selected AME structures and kanamycin A.** *(A)* Kanamycin A is shown as sticks and colored according to the atom type: C — green or magenta, N — blue, O — red, H — white; the dashed circles mark the substituents modified by selected AMEs: *(B)* AAC(6′)-Ib, *(C)* ANT(4′), and *(D)* APH(3′,5″)-IIIa. In panels *(B–D)*, the proteins are presented in cartoon, the ligands are represented as van der Waals spheres: kanamycin A in green, ATP (or ACO in the case of AAC) in dark blue. The ANT structure is composed of two symmetrical parts and only one is colored.

AAC(6′)-Ib (Figure 1.6B) has a unique sequence among acetyltransferases, but its genetic location within mobile elements makes it the most widespread enzyme of the AACs.[55,59] While almost all aminoglycosides can form complexes with AAC(6′)-Ib, only those possessing the 6′-NH_2 group can be modified,[60] such as neamine, kanamycin A or amikacin in Figure 1.1, but not paromomycin. Through site-directed mutagenesis of this enzyme, the amino acids crucial for its enzymatic activity and folding were identified.[61–64] Structurally, AAC(6′)-Ib resembles other enzymes from the family,[60,65] but it differs in that it is active only as a monomer.

ANT(4′)-I (Figure 1.6C) is a promiscuous enzyme, conferring resistance to many aminoglycosides[55] and accepting different cofactors.[66] The available information for this enzyme is rather scarce. It has been shown that the nucleotide binds after an aminoglycoside,[67] which is a unique feature among all AMEs. The two X-ray structures (ANT(4′)-I without[66] and with the substrates[68]) are the only structural data available for the enzymes from this family. Apart from those data, NMR experiments showed that this enzyme is relatively rigid and thus some aminoglycosides have to adjust their conformation upon binding.[67]

The last enzyme, APH(3′,5″)-IIIa (Figure 1.6D), is commonly found in Gram-positive bacteria.[55] Experimental mutagenesis studies[58,69–71] highlighted the amino acids that are the most important for the reaction it catalyzes. It has been shown that the drug binds after

ATP and magnesium ions.[72] By comparing available APH(3′,5″)-IIIa X-ray structures, the conformational changes of the enzyme before and after complexation with an aminoglycoside could be studied,[73–75] and the binding poses of different antibiotics compared.[76,77]

Through a comparison of the available static structures of AME:AG complexes,[67,78–80] some aminoglycosides were found to be dynamic inside the binding sites, while others changed their conformation upon binding, but the majority were found to adopt the same conformation as in the ribosomal A-site. This finding, unfortunately, hampers the process of designing modified aminoglycosides that would have a higher affinity towards the RNA binding site than towards AMEs.

1.2 Computational studies of aminoglycosides and their binding sites

Aminoglycoside antibiotics and their biological partner, RNA, have been a subject of many different computational studies. One of the first simulations concerning aminoglycosides was conducted by T. Hermann and E. Westhof,[81] who showed that when these drugs bind to a hammerhead ribozyme, the positions of their amine groups match those of the Mg^{2+} cations that are bound to the ribozyme during a normal cleavage. These cations are required for proper ribozyme folding and functioning, thus when the antibiotic is bound instead, the ribozyme is no longer active.

Concerning the primary binding site of aminoglycosides, the rRNA A-site, Réblova *et al.* analyzed the internal motion of helix 44 from 16S rRNA, which contains the A-site.[82] These all-atom MD studies showed that adenines A1492 and A1493 are flexible, alternating between the flipped-in and flipped-out states. There was also a transient hydrogen bonding between A1492 and A1408, when A1492 occupied the flipped-in conformation. Interestingly, only base 1492 was observed in the flipped-out conformation. A more complete view of the adenine flipping was presented in a study by K. Y. Sanbonmatsu[83] that used an enhanced sampling method, called Replica Exchange MD, to demonstrate that adenines, A1492 and A1493, are in a dynamic equilibrium between the flipped-in and flipped-out states, which agreed with previous experiments.[18,19] These simulations also suggested a lower energy barrier for the transition of A1492 than for A1493. Another MD simulation study compared the adenine mobility between the free A-site and a complex with an aminoglycoside.[20] The researchers proposed an explanation for aminoglycoside action, where the binding of aminoglycoside narrows the range of movement of A1492 and A1493, rendering these bases incapable of functioning as the differentiating factor between cognate and non-cognate tRNAs.

Several all-atom molecular dynamics (MD) studies of aminoglycosides complexed with the A-site have been performed. Vaiana *et al.* showed that the A-site:PAR complex is stable and the interactions between the antibiotic and RNA are mostly electrostatic, but water-mediated hydrogen bonds were also found to be important.[84] Our previous studies expanded these findings by comparing the A-site internal dynamics of wild-type and eukaryotic-like structures (i.e., possessing the A1408G mutation).[85] Results from our MD simulations showed that this mutation caused a change in the mobility of A1492 and A1493 and also affected the electrostatic potential inside the binding cleft. Another study of the A-site complexes with different aminoglycosides revealed a correlation between the hydration pattern around the bound drug and their computed binding affinity.[86] This work also described how the dynamics of the U1406·U1495 base pair is affected by various antibiotics. Finally, Replica Exchange MD was used to study binding and unbinding of an aminoglycoside to and from the A-site.[87] Results from this study concluded that the steering mechanism

for the RNA:AG association is stochastic gating rather than induced fit, as was proposed earlier. This means that it is not the antibiotic that pushes the adenines A1492 and A1493 to the flipped-out state. Rather, the aminoglycoside has to wait for this state to occur in order to bind.

A more coarse-grained approach was used to investigate the association pathways for aminoglycosides approaching the A-site model[88] and the whole 30S ribosomal subunit.[89] The results from these Brownian dynamics simulations described the driving forces of this association, wherein aminoglycosides appear not to find their 30S binding site immediately but instead they scan the surface for the right match. During the final steps of complex formation, the rate of association depends on both, electrostatic and steric interactions.

Finally, a binding free energy analysis has been reported[90] that combined experimental and computational methods. A detailed electrostatic analysis of both, the A-site model and the whole 30S subunit, showed that the existing models and methods that calculate the electrostatic contribution to the binding free energy yield reliable results, however, specific adjustments to the parameters are sometimes necessary.

In the case of AMEs, only two computational studies have been reported: of AAC(2′)-Ic[91] and AAC(3)-IIIb.[92] This research, however, focused not on aminoglycosides but on the cofactor binding and dynamic response of the enzymes.

1.3 Motivation of the research

The Infectious Diseases Society of America published a new call in 2010, encouraging organizations and scientists to develop ten new antibiotics by the year 2020.[93] In 2011, World Health Day, organized by WHO, was dedicated to drug resistance*. These two organizations, among others, pointed out the diminishing focus of industry on antibacterial innovation, which resulted in only a few new antibiotics introduced to the market in 2010, with most being slight modifications of older drugs. One reason for this decreased output is the cost of developing a new drug, which is estimated to be on average several billions USD.[94] However, another reason is the poor understanding of bacterial resistance mechanisms. Some general schemes are known, but their details often remain unclear. The scientific project presented here seeks to describe the detailed mechanisms of bacterial resistance against aminoglycosides and can help speed the development of better antibiotics from this class.

In the case of aminoglycosides, the resistance methods developed by bacteria are not new. To date, many modified aminoglycosides have been proposed in the hope that they would have higher potency against resistant bacteria.[95] Modifications to the drugs include addition and/or substitution of chemical groups in known aminoglycosides (e.g., Refs.[96–101]), fusion of aminoglycosides with other antibiotics (e.g., Ref.[102]) or even *de novo* re-design of compounds having similar features.[29] However, the majority of these modified drugs, even if they are active *in vitro*, fail to pass the *in vivo* tests, which emphasizes the need to extend our knowledge on the processes occurring inside bacterial cells.

The above mentioned experimental studies on the A-site mutations revealed the correlation between specific base substitutions and effectiveness of different aminoglycosides.[45–47] However, these studies could not explain the underlying physical and chemical changes that led to such an effect. Thus, the question can be posed: how does a change in only one base substantially alter the susceptibility profile of a bacterium? Using molecular modeling of the mutated 16S rRNA A-site fragments and their dynamics, we searched for changes that could explain this phenomenon.

*www.who.int/world-health-day/2011/en/index.html

Furthermore, while there have been extensive studies of aminoglycoside modifying enzymes in terms of kinetics and the thermodynamics of the chemical reactions they catalyze,[14,55] the general lack of structural data for AMEs coupled with their diversity, makes computational studies of these enzymes difficult. This structural, functional and sequence-related diversity of AMEs is in stark contrast to their narrow specificity towards the aminoglycosides, which are a homologous group of antibiotics. This contradiction was the driving force of the second part of the research presented here, which compares representative AMEs according to the way they interact with aminoglycosides and seeks to identify similarities within their binding site features.

1.4 Research aims and tools

The main aim of the thesis is to compare and contrast physical features of the different aminoglycoside binding sites to shed light onto aminoglycoside resistance mechanisms at an atomic level. We have used a biophysical technique called molecular dynamics (MD) to investigate the dynamics of aminoglycosides and their complexes with their pharmacologically-relevant primary binding site, 16S rRNA A-site model, and AME representatives: AAC(6′)-Ib, APH(3′,5″)-IIIa and ANT(4′)-I.

The all-atom simulation of the prokaryotic A-site model enabled us to study the internal movements of this rRNA binding site, and specifically the bases that are responsible for the fidelity of translation and aminoglycoside binding. Next, we conducted simulations of the A-site model with specific nucleotide mutations, to look for changes in the physicochemical properties that these mutations introduce. The simulations of the bare mutated A-site models were complemented with the simulations of their complexes with aminoglycosides. Thus, our computational analysis could be compared with the experiments that provide data on bacterial resistance due to A-site mutations. Finally, we studied the dynamics of the AME representatives and compared the interactions formed with aminoglycosides in these different enzymes. Moreover, we related these interactions to those found in the rRNA binding site.

We used the MD simulations to characterize the dynamics of the selected biomolecules that bind aminoglycosides. There are many ways of analyzing MD simulations, and sometimes different tools must be used to examine proteins and nucleic acids. We have investigated different physical and chemical properties of the simulated systems: hydrogen bonding networks, residue fluctuations, patterns of internal motions, density of water, shape of binding pockets and energetics of aminoglycoside binding. MD simulations were complemented with other biophysical techniques such as the Poisson-Boltzmann model to investigate the electrostatics of the systems, principal component analysis (PCA) to identify differences in biomolecular conformations, and molecular mechanics (MM) generalized Born (GB) solvent accessible (SA) approximation of the binding free energies of the complexes. In addition, we developed a new method that identifies dynamic domains based on a set of molecular conformations and helps describe the conformational changes occurring in biomolecules.

Chapter 2

Methodology

We studied the dynamics of aminoglycosides and selected biomolecules that bind them inside a bacterial cell. Moreover, we investigated and compared the interactions that were formed in the complexes of these different biomolecules with aminoglycosides. This aim required use of a variety of computational tools. We employed MD simulations to study the internal motions of the molecules, but we also analyzed static X-ray structures with the use of Poisson-Boltzmann continuum electrostatic theory. The parameterization of the aminoglycoside molecules required application of quantum calculations. We used several methods to analyze the output of MD simulations — some were already implemented in the available software and for some types of analyses we prepared our own scripts. In addition, we also developed a new algorithm and software to study the internal motions of biomolecules based on their conformations. This chapter presents the theory behind our research. The software we used is listed in the next chapter, section 3.4 *Data analysis*, and the newly developed software is described in Chapter 4, section 4.1 *New method for identifying similarly moving atoms in molecular conformation sets*.

2.1 Quantum calculations of electrostatic potential

In order to create an atomic model of a biomolecule, first one needs a set of coordinates for each atom. Typically, this is provided by experiments, X-ray analysis of crystallized biomolecules or nuclear magnetic resonance (NMR) studies in solid-state or solutions.[103] The Protein Data Bank (PDB)[104] stores the results of these experiments and provides a uniform description of biomolecular conformations.

The atomic positions are not enough, though, if one wants to study how the atoms *interact* with each other. Here comes in physics. The proper, quantum mechanical (QM) approach requires taking into consideration interactions between each pair of the subatomic particles in each atom, which is enclosed in the Schrödinger's equation (Chapter 2 in Ref.[105]):

$$\hat{H}\Psi_{(\vec{R},\vec{r})} = E\Psi_{(\vec{R},\vec{r})} \,, \tag{2.1}$$

where $\hat{H}$ denotes the full Hamiltonian of the studied system (i.e., the operator combining potential and kinetic energies); $\Psi_{(\vec{R},\vec{r})}$ is the wave function, dependent on positions of the nuclei ($\vec{R}$) and of the electrons ($\vec{r}$); and E is the total energy of the system.

Solving the Schrödinger's equation is, however, infeasible even for a powerful computer for any biologically relevant biomolecule, which contains several tens or hundreds of atoms. However, one can notice that a nucleus of an atom moves much more slowly than electrons, due to its larger mass. Thus, when considering nuclei as static particles, only

electron wave functions need to be calculated — this forms the Born-Oppenheimer approximation (see e.g., Chapter 6 in Ref.[105] or Chapter 9 and 10 in Ref.[106]):

$$\hat{H}^{el}_{(\vec{r};\vec{R})}\,\Psi^{el}_{(\vec{r};\vec{R})} = E^{el}\Psi^{el}_{(\vec{r};\vec{R})}\ , \tag{2.2}$$

which describes the movement of electrons ($\vec{r}$) for a given static configuration of nuclei ($\vec{R}$). This assumption is valid only for atoms in their electronic ground state. The wave function for an electron (Ψ^{el}) is dependent on the coordinates and spin value. The part that is controlled by the coordinates is called the atomic orbital (AO).

Having described one atom, a problem appears when one wants to describe a molecule, i.e., a set of atoms that are connected in a precisely defined way through chemical bonds. An approach used here is called the molecular orbital theory (MO; e.g., Chapter 8 in Ref.[105] or Ref.[107]), where the electrons originating from each atom are spread and mixed within the molecule, so that each of them contributes to each bond. This enables to ignore the relativistic effects and to present the wave function of a molecule as a linear combination of the wave functions of its atoms (LCAO, linear combination of atomic orbitals):

$$\Psi^{mol} = \sum_{a} c_a\ \Psi_a\ , \tag{2.3}$$

where index a traverses all atomic orbitals within the molecule; and c_a are numerical coefficients.

Thus, to describe a molecule, the coefficients c_a have to be found. This is obtained through the self-consistent field method (e.g., Chapter 8 in Ref.[105] or Ref.[107]), where an initial guess is successively corrected until another iteration does not change the total energy of the molecule. In this method, for the description of the atomic orbitals, one uses base functions, commonly Gaussian-type functions. The standard basis sets are built from linear combinations of Gaussian-type orbitals (GTO). For example, the basis 6–31G* [108] describes the orbital closest to the nucleus, with a linear combination of six GTOs (6-31G*); next, the other electrons are described by two combinations of three and one GTO, respectively (6–31G*); additionally, the asterisk (6–31G*) denotes that more accurate calculations are provided for heavy atoms, where higher orbitals are included through the so-called polarization functions.[109] The mentioned basis set has been demonstrated to perform excellent for aqueous solutions of biomolecules.[110]

Having found Ψ^{mol}, one can calculate the total energy, E^{mol}, from Equation 2.2, and then the electrostatic potential (ESP, Φ) around a molecule:

$$E^{mol} = -\mathrm{grad}\ \Phi\ . \tag{2.4}$$

The obtained ESP can then be used for assigning partial charges to each modeled atom that best reflect the QM-derived ESP. First, the guessed charges are used to calculate the ESP, which is then fitted, through the least-square fitting procedure, on the correct QM ESP. This method is called restrained electrostatic potential method (RESP).[110] Sometimes a simpler method is used, where charges are assigned to atoms based on precalculated values stored in database. The charges are chosen to match the atom and bond type. One of the best models is called AM1-BCC,[111,112] which gives charges comparable to those obtained through quantum calculations on the 6–31G* basis set.[113,114]

2.2 Molecular mechanics and force field

For larger molecules, it is more convenient to ignore the discrimination between the movement of electrons and nuclei, since electrons can adapt to the movement of nuclei very fast. Within the Born-Oppenheimer approximation, we can describe a molecule as a set of classical particles, with certain radii and point charges, that interact with each other. With such an approximation, one can combine all the "observable" chemical interactions into an equation describing the potential energy of a given molecular configuration, called the force field (FF; e.g., Chapter 7 in Ref.[105] or Chapter 8 in Ref.[115]):

$$V_{FF} = \underbrace{V_{bond} + V_{angle} + V_{tors} + V_{impr}}_{\text{bonded interactions}} + \underbrace{V_{el} + V_{vdW}}_{\text{non-bonded interactions}} . \tag{2.5}$$

The bonded interactions are usually approximated by harmonic potentials:

$$V_{bond} = \sum_{1-2} K_b (b - b_0)^2 , \tag{2.6}$$

$$V_{angle} = \sum_{1-2-3} K_\Theta (\Theta - \Theta_0)^2 , \tag{2.7}$$

$$V_{tors} = \sum_{1-2-3-4} K_\chi [1 + cos(n \cdot (\chi - \chi_0))] , \tag{2.8}$$

$$V_{impr} = \sum_{1-2-3-4} K_\psi (\psi - \psi_0)^2 , \tag{2.9}$$

while the non-bonded terms are:

$$V_{el} = \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon\epsilon_0 r_{ij}} , \tag{2.10}$$

$$V_{vdW} = \sum_{i,j} \varepsilon_{ij} \left[\left(\frac{R^{ij}_{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R^{ij}_{min}}{r_{ij}} \right)^{6} \right] , \tag{2.11}$$

where r_{ij} is the distance between atoms i and j; ε_{ij} is a well-depth, which indicates of a magnitude of favorable interactions between atoms i and j; and the parameters include:

- force constants for: the bonds (K_b [kcal/(mol·Å^2)]), angles (K_Θ [kcal/(mol·rad^2)]), torsional angles (K_χ [kcal/mol]) and the improper torsional angles (K_ψ [kcal/(mol·rad^2)]);
- equilibrium constants for: the bonds (b_0), angles (Θ_0), torsional angles (χ_0; see Figure 2.1A) and improper torsional angles (ψ_0; Figure 2.1B);
- the periodicity parameter (n);
- partial atomic charges (q_i, q_j);
- equilibrium distances between two atoms that are not connected through a chemical bond (R^{ij}_{min}; Figure 2.1A).

All the parameters are given values based on experiments (spectroscopy or NMR) and quantum calculations (e.g., Ref.[116] or Chapter 8 in Ref.[115]), and they depend on the types of the atoms that form the interaction.

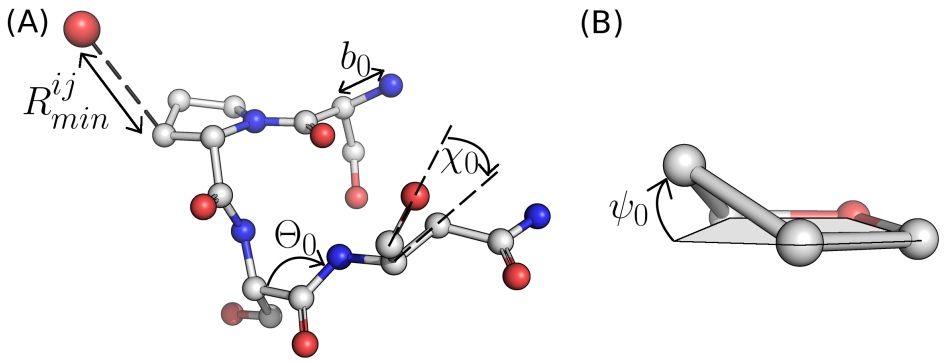


Figure 2.1: **Illustration of the force field components**. See the text for description.

2.2.1 Energy minimization

The potential energy of a molecule depends on its conformation, i.e., the positions of all atoms. As has been mentioned, these initial positions used in simulations are typically obtained from experiments. However, the X-ray derived structures lack coordinates for hydrogen atoms, and therefore these have to be added *in silico*. Moreover, usually one adds also water and ions, to reproduce the natural environment of a molecule. This addition can result in some atoms being too close to each other. Therefore, a minimization of energy is conducted prior to any other calculations. During the minimization, the atoms are being slightly repositioned so as to reach the minimal potential energy of the whole system.

There are several algorithms for finding this minimum — in our simulations, two of them were used, namely the steepest descent (SD) and conjugate gradients (CG) algorithms. These fall into the category of linear descent search algorithms (see Chapter 10 in Ref.[115]), for which a general scheme is outlined below.

General linear descent search algorithm:

1. at a given point $\vec{x}_i$ of the potential energy surface (PES) calculate the energy $E(\vec{x}_i)$ and its gradient $\vec{g}_i = \mathrm{grad}\, E(\vec{x}_i)$;

2. create a search vector $\vec{s}_i$ (dependent on the type of algorithm);

3. move on PES along $\vec{s}$:
$$\vec{x}_{i+1} = \vec{x}_i + \lambda_i \vec{s}_i \; ; \tag{2.12}$$

4. calculate the new energy $E(\vec{x}_{i+1})$ and the energy difference: $\Delta E = E(\vec{x}_{i+1}) - E(\vec{x}_i)$;

5. repeat the above steps until $\Delta E < \varepsilon$.

In the above scheme, both λ_i and ε are very small but non-zero real numbers. In order to obtain the best λ_i value, the line-search algorithm is used. Using Equation 2.12, three different movements (for three different λ_i values) along $\vec{s}_i$ are generated, and the energy in these new points is calculated. Then, a polynomial is fitted to this set of points $E(\lambda_i)$. The minimum of this polynomial gives the best λ_i parameter for a given step.

SD and CG algorithms differ only in the definition of the search vector $\vec{s}$. In SD it is based only on the current point in PES: $\vec{s}_i^{\,SD} = -\vec{g}_i / |\vec{g}_i|$; while the CG version depends on the previously chosen search vector and the current gradient: $\vec{s}_i^{\,CG} = -\vec{g}_i / |\vec{g}_i| + \beta_i \, \vec{s}_{i-1}^{\,CG}$, where β_i is a constant dependent on the current and/or previous gradient vector, and it is specific to each implementation of the CG algorithm.[115]

The SD algorithm performs better when the search for the minimum is in the initial stage, but when being already close to the minimum, this approach can result even in stepping over the minimum and continuing to the other slope of PES. This is undesirable, and therefore, commonly, the minimization is started with SD and continued with CG, because the latter algorithm performs better near the minima.[117]

2.3 Molecular dynamics

2.3.1 Theory behind

With such a representation of the biomolecule and the interactions between its atoms, one can now proceed to calculate the forces:

$$\vec{F} = -\text{grad}\left(V_{(\vec{x})}\right) \ , \tag{2.13}$$

which would lead to movement, according to the Newton's equation:

$$\vec{F} = m \cdot \ddot{\vec{x}} \ . \tag{2.14}$$

The analytical solution of the above equation, however, includes integrals, which are difficult to implement in a computer, thus several approximate algorithms have been developed. The positions in consequent time steps, $\vec{x}(t-\Delta t)$ and $\vec{x}(t+\Delta t)$, are expressed through Taylor approximations truncated to the forth order terms (e.g., Ref.[118] or Chapter 7 in Ref.[105]):

$$\vec{x}(t+\Delta t) \overset{Taylor}{=} \vec{x}(t) + \Delta t\ \vec{v}(t) + \frac{1}{2}\Delta t^2\ \vec{a}(t) + \frac{1}{6}\Delta t^3 + \mathcal{O}(\Delta t^4) \ ,$$
$$\vec{x}(t-\Delta t) \overset{Taylor}{=} \vec{x}(t) - \Delta t\ \vec{v}(t) + \frac{1}{2}\Delta t^2\ \vec{a}(t) - \frac{1}{6}\Delta t^3 + \mathcal{O}(\Delta t^4) \ .$$

Addition and subtraction of the above equations yields, respectively:

$$\vec{x}(t+\Delta t) = 2\vec{x}(t) - \vec{x}(t-\Delta t) + \vec{a}(t)\,\Delta t^2 + \mathcal{O}(\Delta t^4) \ ,$$
$$\vec{v}(t) = \frac{\vec{x}(t+\Delta t) - \vec{x}(t-\Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2) \ ,$$

where $\vec{x}$, $\vec{v} \equiv \dot{\vec{x}}$ are the position and velocity of an atom; Δt is the timestep; and $\vec{a} \equiv \ddot{\vec{x}}$ denotes acceleration, which can be obtained from Equation 2.14. This forms the so-called Verlet algorithm,[118] which is very efficient and gives accurate positions of atoms ($\mathcal{O}(\Delta t^4)$). However, one of the disadvantages of the Verlet algorithm is that the velocities are computed with relatively large errors ($\mathcal{O}(\Delta t^2)$), and therefore sometimes another approach is used, called leap-frog,[118] which provides better estimation of the velocities. In the leap-frog algorithm, an additional mid-point of the timestep is created to facilitate the evaluation of the mutually dependent positions and velocities:

$$\vec{x}(t+\Delta t) = \vec{x}(t) + \vec{v}(t+1/2\Delta t)\,\Delta t \ ,$$
$$\vec{v}(t+1/2\Delta t) = \vec{v}(t-1/2\Delta t) + \vec{a}(t)\,\Delta t.$$

As mentioned previously, the initial positions are given, but the initial velocities have to be somehow assigned. Most often, these are randomized values from the standard Maxwellian velocity distribution, which are generated based on a number created from the current date and time, in order to minimize the probability of repetition.[118] The sequence of atomic positions, velocities, and accelerations yields an MD trajectory.

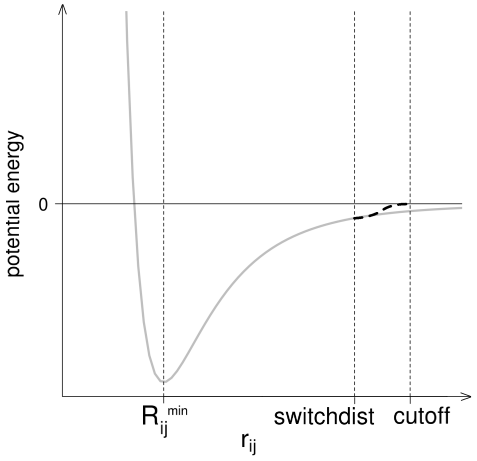


Figure 2.2: **Exemplary function for the Lennard-Jones-type van der Waals interactions** (solid light grey line): when the atoms i and j are approaching each other, the energy becomes attractive, until the atoms are within the optimal distance (R^{ij}_{min}), from which the repulsive term is the dominant one. The switching function (dashed line), which can be turned on at `switchdist` point, helps to smooth the switch of the potential energy to zero for distances larger than the `cutoff` point.

In order to precisely calculate the movement, one needs to have a good approximation of the conformation-dependent potential energy from the force field. The only difficulty is the calculation of the non-bonded interactions, which in principle should be computed for each pair of the atoms, regardless of their distance. This would bring, however, huge computational costs when simulating larger biomolecular systems. One can see from Equations 2.10 and 2.11 that these non-bonded terms are small enough at large distances to be neglected without much loss of precision. The van der Waals interactions decline as fast as r^{-6}, and therefore the standard approach for computing these interactions is to set a `cutoff` distance parameter — above this value, this non-bonded term is set to zero. To allow a smooth transition, one can also utilize a switching function from a set `switchdist` distance to the `cutoff` (Figure 2.2). When the two parameters are set reasonably, the calculations are faster and the precision is maintained.

The electrostatic interactions diminish much slower than the van der Waals terms (r^{-1}), and therefore, typically, a different approximation is used, called particle mesh Ewald (PME) summation.[119] The short-range electrostatic interactions are computed explicitly, through Equation 2.10, in the Euclidean space, while the long-range ones are calculated in the Fourier space:

$$V_{el} = \sum_{i,j} \phi(\vec{x}_j - \vec{x}_i) = V_{short} - V_{long} \; ; \tag{2.15a}$$

$$V_{short} = \sum_{i,j} \phi_{short}(\vec{x}_j - \vec{x}_i) \; , \tag{2.15b}$$

$$V_{long} = \sum_{\vec{k}} \tilde{\phi}_{long}(\vec{k}) \left| \tilde{\rho}(\vec{k}) \right|^2 \; , \tag{2.15c}$$

where ϕ is the electrostatic potential, and $\tilde{\phi}$, $\tilde{\rho}$ are the Fourier transforms of the electrostatic potential and of the charge density, respectively. Each of these functions, V_{short} and V_{long}, quickly converges to zero in their own space, thus it is sufficient to take into account only a couple of first terms when calculating these interactions.

As with all physical models of reality, the problem of boundary conditions exists. Ideally, our biomolecule of interest would be surrounded by an infinite amount of solvent (we do not consider here models that do not require solvent, nor simulations of membranes, which form a natural boundary). The infinity is, of course, nonreproducible, and therefore, in the classical MD simulations, the entire finite box with the solvent and the solute is copied — an approach called periodic boundary conditions (PBC; illustrated in Figure 2.3; see e.g.,

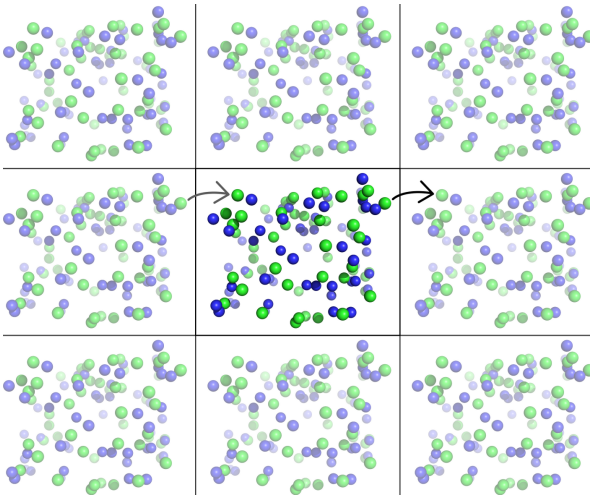


Figure 2.3: **Periodic boundary conditions.** The simulated system is finite (the cell in the middle) but the coordinate system is "wrapped" so that each particle moving beyond the border of the system appears at the opposite side with the same velocity.

Chapter 12 in Ref.[115]). Hence, all the atoms are surrounded by other atoms, which prevents the atoms near the edge from "escaping", and if an atom moves towards the edge of a box, it would simply appear at an opposite edge, moving with the same velocity. In practice, the minimum image convention is used, where the trajectory of only one copy of the system (N atoms) is recorded, and each atom is allowed to interact only with the other $N - 1$ atoms that are in the closest copy.

2.3.2 Simulation setup

The MD simulation protocol is composed of several steps:

minimization, described in section 2.2.1;

thermalization, when the temperature of the simulated box is slowly raised to a desired one (commonly to the room temperature, 300–310 K);

equilibration, this step lets the molecule and the solvent "get used to" the higher temperature and it is the last preparation step;

production, where the main data is gathered for a subsequent analysis; this step is the longest one — depending on the size of the system and available computational power, it can last from tens to thousands of nanoseconds.

Importantly, the above list is only a scheme of the process — for each of the simulated biomolecule, one has to adjust this scheme in order for the simulations to yield reliable and reasonable results.

Thermalization is required because the initial coordinates describe molecules at a different temperature — the one that had been used in the experiment (e.g., typically $\sim$ 100 K for the X-ray studies[120]); while the simulations are usually conducted at room temperature ($\sim$ 300 K). Equilibration is also necessary, since in this step the density of the solvent is being leveled so that the whole simulated box is filled equally. Moreover, the conformation of a molecule obtained from X-ray studies can be slightly different than its conformation in the physiological conditions due to different forces in the crystal (crystal packing artifacts[120]).

During the production stage, when no specific constraints are imposed on the simulation conditions, the simulation revolves in the microcanonical ensemble,[118] also called NVE because of the three constants: the number of atoms (N), the volume (V) and the energy (E). This is a natural setup but some systems or aims require a different ensemble. Moreover,

the numerical errors and implementation issues may result in a rise of temperature and/or energy if the simulation parameters were not adjusted to a given system. In this research, the NVE ensemble was used for relatively short simulations of compact proteins. The elongated RNA fragment was, however, simulated in the NPT ensemble. Because of its higher flexibility and being a fragment of a much bigger construct, it was more natural to apply the constant pressure (P) and temperature (T) in this system. A simulation in the NPT ensemble is also termed Langevin dynamics (see Chapter 13 in Ref.[115] or Ref.[121]), due to the Langevin equation used to describe the forces:

$$m \cdot \ddot{\vec{x}}(t) = -\Delta V_{(\vec{x}(t))} - \gamma\, m\, \dot{\vec{x}}(t) + \vec{F}_r(t)\ , \tag{2.16}$$

where γ is the collision parameter $[1/s]$, which defines the friction; and $\vec{F}_r(t)$ denotes an additional random force that is generated from a Gaussian distribution:

$$\left\langle \vec{F}_r(t) \right\rangle = 0\ ; \tag{2.17a}$$

$$\left\langle \vec{F}_r(t)\vec{F}_r(t=0) \right\rangle = 2k_B T_0\, m\, \gamma\, \delta(t)\ , \tag{2.17b}$$

where k_B is the Boltzmann constant and $\delta(t)$ is the Dirac delta function. These stochastic additional forces mimic collisions between the solute and solvent molecules, which, as can be seen in Equation 2.17b, provides a coupling to a heat bath with a set temperature of T_0.

The timestep in the simulations has to be large enough to obtain long trajectories that would allow to observe biologically relevant motions, but short enough to reproduce all the interactions correctly. Therefore, the optimal step is connected to the most vibrating bonds, i.e., bonds with hydrogen atoms. These are movements lasting approximately 10 fs (frequencies of around 3000 $1/\mathrm{cm}$ [122]). Therefore, usually the timestep is set to a one order higher value of 1 fs. This can be extended to 2 fs when putting harmonic restraints on these bonds formed with hydrogen atoms, as it is e.g., in the SHAKE algorithm.[123]

2.4 Basic molecular dynamics data analysis

MD simulations produce nowadays a lot of data that has to be carefully analyzed. The most basic quantities are the temperature and the energy of the simulated system. In the classical MD these values should stay leveled in the production phase; therefore, the first check of the stability of the simulation is to visualize the time dependence of these values. The energy and temperature would naturally rise during the thermalization and equilibration but should hit a plateau in the production stage.

However, the most important data obtained from MD simulations are the positions of atoms versus time. These constitute sometimes even hundreds of thousands of conformations of a studied molecule or complex. There are many ways one can analyze these data, depending on what the aim of the research is and what types of approximations were used in the model. However, the first and basic analysis is to observe whether the molecule is internally stable. Unless one performs a simulation of folding or unfolding, the global conformation of the structure should be maintained if the simulation was prepared correctly. The standard procedure is to calculate the global root-mean-square deviation (RMSD) of atomic positions from the reference structure, which is commonly the starting one.

$$\mathrm{RMSD}(t) = \sqrt{\frac{1}{M}\sum_{i=1}^{n} m_i \left|\vec{x}_i(t) - \vec{x}_i(0)\right|^2}\ , \tag{2.18}$$

where $M = \Sigma_{i=1}^{n} m_i$; m_i denotes the mass of atom i; n is the total number of atoms in the molecule; $\vec{x}_i(t)$ shows the position of atom i in timestep t; and $\vec{x}_i(0)$ points to the starting position of atom i.

Another measure of compactness of the structure is the radius of gyration, R_{gyr}:

$$R_{\mathrm{gyr}}(t) = \sqrt{\frac{M \left|\vec{x}_i(t) - \vec{x}_{\mathrm{COM}(t)}\right|^2}{M}} , \tag{2.19}$$

where $\vec{x}_{\mathrm{COM}}(t)$ is the position of center of mass (COM) at timestep t. This gives a rough estimate of the size of the sphere that would contain the given structure, and therefore, the bigger the radius, the less compact the conformation.

Naturally, in the initial stages of the simulation, i.e., minimization, thermalization and equilibration, RMSD and R_{gyr} should rise because the structure changes along with changes in the surrounding environment (raising temperature, interactions with added solvent or ion molecules, etc.). However, in the production stage these values should rather oscillate around a certain mean value, which indicates that the molecule maintains its native, folded state.

There might be, however, differences in the mobility of various fragments of a biomolecule. Another standard measure, root-mean-square fluctuations (RMSF), shows how much each of the atoms or residues (i.e., amino acid or nucleotide) moves in the simulation, with regard to its average position:

$$\mathrm{RMSF}(i) = \sqrt{\frac{1}{T}\sum_{t=1}^{T} \left|\vec{x}_i(t) - \langle\vec{x}_i\rangle\right|^2} , \tag{2.20}$$

where i enumerates the residues or atoms; T is the total simulation time; and $\langle\vec{x}_i\rangle$ denotes the mean position of atom i. These values can be compared with the experimental β-factors (temperature factors [Å^2]), which are parameters used by crystallographers to refine the structure (e.g., Chapter 7 in Ref.[120]). A β-factor tells us how much the atom in the crystal fluctuates isotropically around its position in the model — if that value is large (more than 50 Å^2), the position of the atom in the model is not well defined, and therefore, one can expect high mobility of this atom in the MD simulation.

2.5 Describing internal motions

The measures described above, although quite crude, can indicate a biologically significant event, e.g., a conformation change or large differences in mobility between biomolecular fragments. The identified event often must be more thoroughly examined using some more advanced tools.

2.5.1 Principal Component Analysis

The raw data from an MD simulation constitute of fluctuating atoms, which can appear chaotic and difficult to quantify. A well known physical method can help to emphasize the main directions of internal collective molecular motions. Principal component analysis (PCA) reduces the dimensionality of the movements and finds their main modes.[124] This is done by, first, calculating the covariance matrix, $\hat{C}$:

$$C_{ij} = \left\langle M_{ii}^{\frac{1}{2}}\Big(x_i - \langle x_i\rangle\Big) M_{jj}^{\frac{1}{2}}\Big(x_j - \langle x_j\rangle\Big)\right\rangle , \tag{2.21}$$

where $\hat{M}$ denotes a diagonal matrix with masses of atoms (in case of weighted covariance matrix) or an identity matrix; x_i is the ith coordinate, out of total $3n$ coordinates (n is the number of atoms); and $\langle x_i \rangle$ is the average value of x_i.

This matrix is then diagonalized, in order to find the internal coordinate system:

$$\hat{R}^T \hat{C} \hat{R} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{3n}) \tag{2.22}$$

where $\hat{R}$ denotes an orthonormal invertible matrix, whose columns are eigenvectors of $\hat{C}$; and λ_i is the ith eigenvalue ($\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{3n}$).

Then, the trajectory can be projected onto the eigenvectors (also called modes or principal components, PCs):

$$\vec{p}(t) = \hat{R}^T \hat{M}^{\frac{1}{2}} \left(\vec{x}(t) - \langle \vec{x} \rangle \right) \tag{2.23}$$

Each mode $p_i(t)$ has an associated eigenvalue λ_i, which reflects the value of mean-square fluctuations (Equation 2.20). The couple of modes with the largest eigenvalues describe the most significant motions.

2.5.2 Conformer plots

This description of motions in terms of internal coordinates can be used to compare different conformations and trajectories of the same molecule. Due to the decreasing order of eigenvalues, the two or three first modes are sufficient to distinguish between some important conformational changes within the structure. Thus, already plotting the relation between PC 1 and PC 2 often gives interesting insights. This is especially useful when there are many conformations and/or trajectories to analyze, since in a plot of PC 1 vs. PC 2 one conformation would be depicted as one dot.

Not only the modes defined by PCA can be plotted like that. In fact, one can define any type of internal coordinates, depending on the studied system. The idea behind is still the same — to reduce the dimensionality of the observed conformational changes.

2.5.3 Correlated motions

Sometimes, one would like to know how exactly the conformations differ between themselves, which parts of a molecule change their positions. One approach is to calculate the dynamic cross-correlation matrix (DCCM, $\hat{C}'$):[125]

$$C'_{ij} = \frac{C_{ij}}{N_{ij}} = \frac{\langle \vec{d}_i \cdot \vec{d}_j \rangle_{MD}}{\sqrt{\langle \vec{d}_i \rangle^2_{MD} \langle \vec{d}_j \rangle^2_{MD}}}, \tag{2.24}$$

where $\hat{C}$ is the covariance matrix from Equation 2.21; $\hat{N}$ denotes the normalizing matrix; vector $\vec{d}_i \equiv \vec{x}_i(t) - \langle \vec{x}_i \rangle$ shows the deviation of atom i from its mean position; and $\langle \quad \rangle_{MD}$ denote an average over the whole trajectory.

The correlation matrix is symmetric (i.e., $C'_{ij} = C'_{ji}$) and $C'_{ij} \in [-1; 1]$. The high positive values describe the correlated motions (i.e., two atoms are moving in the same direction at the same time), while values close to -1 point to anti-correlated motions. However, it can be shown that this method loses some information, namely, it does not accounts for the correlated rotations because of averaging over the Cartesian space. Let us consider two points, $a = r_a(\cos\alpha, \sin\alpha)$ and $b = r_b(\cos\beta, \sin\beta)$, that rotate around a common point (Figure 2.4). Their trajectories can be described by the distances from the center (r_a and r_b) and the angles (ϕ_i): $a = r_a(\cos(\alpha + \phi_i), \sin(\alpha + \phi_i))$ and $b = r_b(\cos(\beta + \phi_i), \sin(\beta + \phi_i))$.

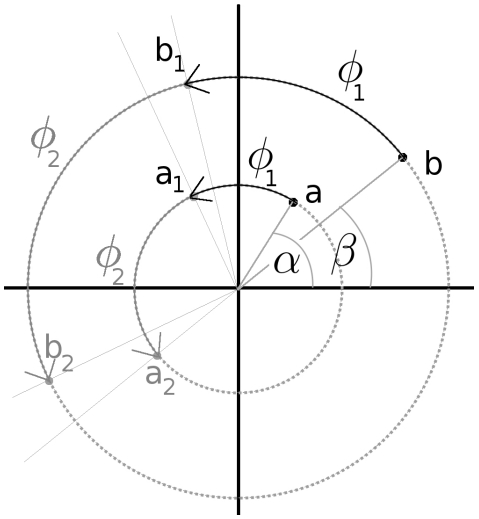


Figure 2.4: **Example of a correlated rotation.** This kind of correlation is difficult to detect with a standard cross-correlation matrix.

Equation 2.24 would then yield a correlation coefficient: $C'_{ab} \sim \cos(\alpha - \beta)$, which therefore cannot be used to state whether the motions are correlated or not. There are, of course, other notations (spherical coordinate system, Euler angles), which would deal better with rotations, but they are much more complicated to implement.

In order to provide a better description of how fragments of the molecule move in relation to one another, we have developed an algorithm and a software, Geometrically Stable Substructures (GeoStaS), which divides a biomolecule into "dynamic domains". Each molecular fragment that appears internally rigid but moves with relation to another fragment, constitutes one dynamic domain. Our algorithm identifies also a correlated rotation shown in Figure 2.4. The details of the new algorithm and its implementation are given in Chapter 4, Section 4.1 *New method for identifying similarly moving atoms in molecular conformation sets*.

2.6 Intra-molecular interactions

When describing the changes either in a conformation of a molecule or in a relative orientation of two molecules forming a complex, the simplest way is to analyze hydrogen bonds that are formed and/or destroyed. Hydrogen bonds can be direct or mediated through a water molecule. The two chemical moieties between which a hydrogen bond can

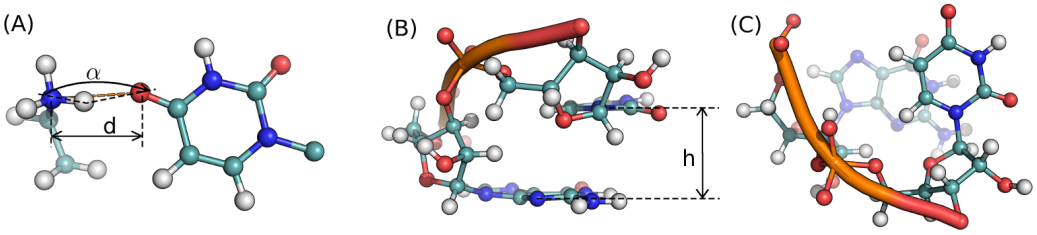


Figure 2.5: **Examples of non-covalent interactions.** *(A)* Hydrogen bonds are formed between the donor (here: oxygen atom, colored red) and the acceptor (here: nitrogen atom, colored blue), through the hydrogen atom (colored silver). The geometrical definition states that the distance d should be less than 4 Å and the angle $\alpha \in [120;\ 180]^\circ$. *(B–C)* Stacking interactions enhance stability and are formed between two aromatic rings — here, two neighboring nucleic bases from a helix are shown *(B)* from the side and *(C)* from top. The optimal distance, h, is ~ 4 Å.

be formed are called a *donor* and an *acceptor* (Figure 2.5A). The donor has at least one free electron pair, while the acceptor has to be able to "accept" an electron, i.e., is covalently bonded to a hydrogen atom. The formal definition of a hydrogen bond[126] is based on a distance between the donor and the acceptor and an angle between the three bond-forming atoms. The closer the angle is to 180°, the stronger the hydrogen bond. While the distance varies depending on the atom type of the donor and acceptor, it is typically less than 4 Å. Thus, monitoring the distance and angle between possible donor–acceptor pairs is a simple way of quantifying hydrogen bond networks during an MD simulation.

Another important non-bonding interaction type is stacking (Figure 2.5B–C), formed between two parallel aromatic rings. In an aromatic ring, the so-called delocalized electrons from π orbitals form an electron cloud over and under the ring. The two clouds from two rings interact with each other through the stabilizing London dispersion forces and repulsive interactions between multipoles. Stacking is very important especially in nucleic acid structures, where the stacked nucleic bases stabilize e.g., the double helix (Chapter 8 in Ref.[127]). In most of the force fields, this effect can be reproduced through the van der Waals interactions. The geometrical definition of the stacking is based on the distance between the two rings, which should stay around 4 Å.[128] The van der Waals attractive force diminishes quite fast for larger distance, and it becomes very quickly repulsive when atoms are closer (see Equation 2.11). When trying to find stacking interactions in a simulated molecule, not only the distance has to be taken into account — one has to check whether the rings are in a parallel orientation. This can be done by constructing vectors ($\vec{a}$ and $\vec{b}$) perpendicular to the planes of the rings (A and B) and calculating a cosine of an angle between them:

$$A||B \Leftrightarrow \cos\left(\measuredangle(\vec{a},\vec{b})\right) \approx \{1;-1\}\ ; \qquad \vec{a}\perp A\ ;\ \ \vec{b}\perp B\ . \tag{2.25}$$

2.7 Electrostatic description

2.7.1 Poisson-Boltzmann model

One of the very basic features of a molecule is its electrostatic potential (ESP). To calculate the ESP for large biomolecules, such as proteins or nucleic acids, one often uses continuum electrostatics theory. This theory is based on the nonlinear Poisson-Boltzmann equation (PB):[129,130]

$$\nabla\cdot[\epsilon(\vec{x})\ \nabla\Phi(\vec{x})] = -4\pi\ \rho_{solute}(\vec{x}) - 4\pi\sum_{i=1}^{n_i} q_i c_i \exp\left(\frac{-q_i\Phi(\vec{x})}{k_B T}\right)\ , \tag{2.26}$$

where ∇ is a divergence operator; $\epsilon(\vec{x})$ describes the position-dependent dielectric function; $\Phi(\vec{x})$ is the ESP at position $\vec{x}$; $\rho_{solute}(\vec{x})$ denotes the charge density of the solute and the Boltzmann distribution $\rho_{total}(\vec{x}) = \sum_{i=1}^{n_i} q_i c_i \exp\left(-q_i\Phi(\vec{x})/k_B T\right)$ is used to describe the solvent charge density, coming from ions (n_i is the number of all ion types, each with charge of q_i); c_i denotes the concentration of ions of type i; k_B is the Boltzmann constant.

Solving Equation 2.26 gives us $\Phi(\vec{x})$, but the analytical solution is not possible for arbitrary shaped objects such as biomolecules. Therefore, several methods have been developed to obtain the numerical solutions of PB equation, including the finite-difference or boundary element methods.[129,131] Sometimes, the full PB equation can be approximated by its linear version, which is obtained when the exponent in the charge distribution in Equation 2.26 is approximated by a Taylor series truncated to the first-order term $\exp\left(-q_i\Phi(\vec{x})/k_B T\right) \approx$

$1 - q_i\Phi(\vec{x})/k_BT$. This holds when the energies are much smaller than the thermal energy, $q_i\Phi(\vec{x}) \ll k_BT$, which can be assumed for dilute solutions or monovalent electrolytes. Then, the electrostatic energy (G_{PB}) of a given biomolecular conformation can be found through a "charging" process:

$$G_{PB} = \frac{1}{2}\int \rho(\vec{x})\ \Phi(\vec{x})\ d\vec{x}\ . \tag{2.27}$$

2.7.2 Generalized Born model

However, even with such an approximation the calculations for hundreds of conformations of a biomolecule can be too time- and resource-consuming to conduct. Therefore, one may choose another approximation, where a molecule is built of many non-overlapping spheres with point charges. This enables to represent the electrostatic solvation free energy through the use of the well-known Born formula for each of the spheres:[131]

$$G_{i,GB} = -\left(1 - \frac{1}{\epsilon_{sol}}\right)\frac{q_i^2}{2a}\ , \tag{2.28}$$

and the total electrostatic energy for the molecule would be:

$$G_{GB} \simeq -\left(1 - \frac{1}{\epsilon_{sol}}\right)\frac{1}{2}\sum_{i,j}\frac{q_iq_j}{f_{ij}^{GB}}\ , \tag{2.29}$$

where ϵ_{sol} is the dielectric constant of the solvent; the summation includes all the "atom" pairs in the molecule; and f_{ij}^{GB} is a function of the distance between atoms i and j, and of the effective Born radii of the atoms, R_i. In the GB approximation this effective radius depends not only on the radius of atom i but also on radii and positions of other atoms. Therefore, when creating the force field parameters for the molecule, one should include these specific effective radii in the parameterization. The main issue with the GB method is its applicability mainly to spherically symmetric systems, due to the formulation of Equation 2.28. Recently, however, a method has been developed that includes a scaling factor for the Born radii of the atoms buried beneath the surface of a molecule.[132] This approach yields the electrostatic energies comparable to those obtained with the PB method but is much faster.

2.8 Assessment of the binding free energy

When analyzing a structure of a complex, it is often desired to know how well the molecules fit to each other, which can be measured by their free energy of binding. This value is especially important when performing calculations associated with the drug design process. Sometimes the two compared complexes may behave similarly in a simulation, but there can be a difference in the binding free energy, which could point to a potentially better drug. One of the methods used with biomolecular simulations to estimate the binding free energies is MMPBSA or MMGBSA, which is named after the division of the total free energy into terms calculated through Molecular Mechanics, PB (or GB), and solvent accessible surface. This method is outlined below.

The binding free energy is naturally defined as a difference between the free energy of the complex and its substrates separately, thus first the free energies of the receptor and the ligand that form the complex have to be calculated. According to thermodynamics, the total energy available in an isolated system (also known as Gibbs free energy, G) is combined of

enthalpy (H) and entropy (S) (see e.g., Chapter 8 in Ref.[133]):

$$G = H - T \cdot S \ . \tag{2.30}$$

The enthalpic part is connected to the internal chemical properties and the energy of interaction with the solvent, thus it can be assessed by the force field energy value (also called molecular mechanics term, V_{FF}) and the solvation energy (G_{PBSA} or G_{GBSA}):[129,134]

$$H = V_{FF} + G_{GBSA} \ . \tag{2.31}$$

The solvation energy is computed through submerging the molecule in an implicit solvent medium, even though the simulations were performed with atomic representation of the water. That is due to implicit models being less computationally demanding and incorporating all the features of the discrete model (such as hydrophobicity or a high dielectric environment) but averaged. Also, one does not need to worry about boundary conditions, since in the calculations the implicit solvent can be assumed infinite. The solvation energy is thus obtained as follows:[129,134]

1. all the molecular charges are zeroed in vacuum;
2. the uncharged solute is solvated;
3. all charges are restored in the implicit solvent environment.

Therefore, the solvation energy can be dissected into:

$$G_{GBSA} = G_{SA} - W_{vac} + W_{sol} \ , \tag{2.32}$$

where G_{SA} and $G_{pol} = -W_{vac} + W_{sol}$ are the non-polar and polar terms, respectively; W_{vac} and W_{sol} denote the work needed for discharging the molecule in vacuum and re-charging it in solvent, respectively.

The polar term is calculated as described in Section 2.7 *Electrostatic description*. The non-polar part above is estimated by the solvent accessible surface area (SASA) of the molecule: $G_{SA} = \gamma \cdot \mathrm{SASA}$, where γ [kcal/(mol·Å^2)] is the surface tension.

The entropic part from Equation 2.30 is associated with the level of order in the structure, i.e., how rigid it is. This term is much more difficult to calculate. There are several approximations, including the quasi-harmonic approximation[135] or based on the normal mode analysis.[136] All these methods, however, become unreliable when a big conformational change upon binding is anticipated. This is the case with aminoglycosides, which have many rotatable bonds, thus their conformation can be significantly altered when forming a complex. In our calculations, after several trials, we decided not to include the entropic term, and therefore the resulting values serve only as a comparison and cannot be treated as absolute.

MD simulation provides multiple conformations, and for each of those the free energy value can be computed and then averaged ($\langle G \rangle$). Then, the final binding free energy is obtained:

$$\Delta G = \langle G_{complex} \rangle - \left(\langle G_{receptor} \rangle + \langle G_{ligand} \rangle \right) \ . \tag{2.33}$$

Although, generally this type of calculation does not yield absolute values, when comparing the resulting energy for different complexes their order is the same as for the experiments (e.g., Refs.[86,90]). Therefore, the MMPB(GB)SA method is often used for a rough estimation of the binding free energy but it is especially valuable in a detailed investigation of the contributions of individual residues (see e.g., Refs.[137–139]). This energy can be easily decomposed into terms that originate from specific residues. This kind of information is helpful when designing modifications of drugs.

Chapter 3

Parameterization and preparation of the simulated systems

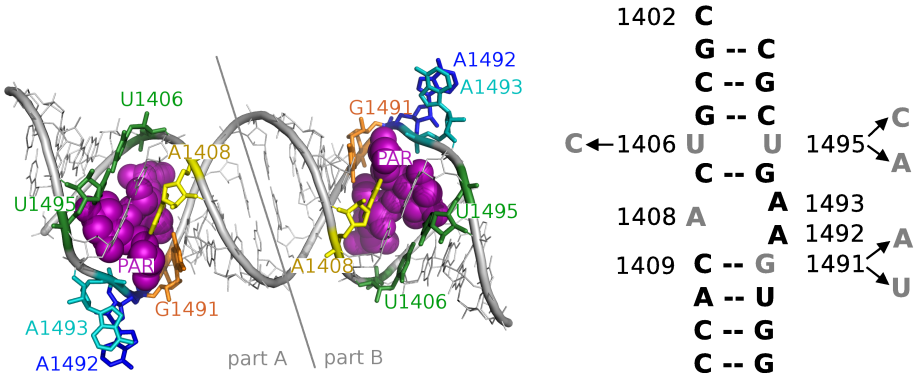
The main aim of this study was to investigate aminoglycoside antibiotics recognition and to provide insight to the underlying mechanisms of bacterial resistance. MD simulations were conducted on a number of systems to study the interactions between aminoglycoside and RNA binding site models, as well as the representative modifying enzymes. Below, the details of the simulation preparation steps are outlined.

3.1 Aminoglycoside binding site in ribosomal RNA

3.1.1 Starting structures

In the cellular environment, aminoglycosides bind to the 30S ribosomal subunit[14] (see Figure 1.3). The entire 30S subunit is a large nucleoprotein complex of about 100 000 atoms[30] and all-atom MD simulations of the entire 30S subunit are extremely computational demanding even at the nanosecond timescale. However, it has been shown that these antibiotics also bind to fragments of this subunit, specifically to the models of the A-site.[140,141] Moreover, the thermodynamic features of this binding process are almost the same as for the binding to the whole 30S subunit.[140,141] One of the A-site models is a symmetrical construct of two aminoglycoside binding sites, thus it can accommodate two drugs (Figure 3.1A), which immediately increases the statistics of observations (we can assume that there is no mutual dependence of these binding sites, since they are divided by eight canonical base pairs).[84,90,141]

We chose the complexes of the model A-site with two aminoglycosides, paromomycin (PAR, PDB ID: 1J7T,[142] resolution of 2.5 Å) and kanamycin A (KAN, PDB ID: 2ESI,[31] resolution of 3.00 Å). These two drugs represent the two classes of aminoglycosides: 4,5- and 4,6-disubstituted 2-DOS (see Figure 1.1B–C). The complex of the A-site with kanamycin A was simulated in order to have a reference for the subsequent simulations of aminoglycoside modifying enzymes (described below). The complex with PAR was subjected to *in silico* mutations (presented in Figure 3.1B). We have chosen those mutations that introduced high level of resistance, especially against paromomycin, as reported in the experiments on whole bacteria (referenced to and described in Table 3.1). Also the A1408G mutation confers high resistance in bacteria (see Table 1.1). We studied this mutation as a separate project and the results have been published previously.[85] Here, we focus on some other base substitutions. All types of simulations concerning the A-site are listed in Table 3.1.



(A) Starting structure for the RNA simulations (B) Sequence of the A-site

Figure 3.1: **The RNA A-site model used in the simulations**: *(A)* The whole model (PDB ID 1J7T) complexed with paromomycin (PAR). *(B)* The sequence of the A-site with marked mutations that were studied here.

Table 3.1: **MD simulations of RNA and the abbreviations used.**

simulation of RNA		*simulation*	
with AG[(a)]	*without AG*	*time [ns]*[(b)]	*effect of introducing mutation*
WT_PAR	**WT**	20	—
G1491A_PAR	**G1491A**	20	A is found in eukaryotic sequence[15,16] and confers resistance against paromomycin (up to 64-fold increase in MIC values[46,48])
G1491U_PAR	**G1491U**	20	high resistance against paromomycin (512-fold increase in MIC[42])
U1495C_PAR	**U1495C**	20	resistance against paromomycin (128-fold increase in MIC in *M. smegmatis*[42] and 5-fold increase in *T. thermophilus*[46])
U1406C/U1495A_PAR	**U1406C/U1495A**	20	high resistance against many aminoglycosides (>1000-fold increase in MIC for paromomycin[42,49])
WT_KAN	—	25	—

[(a)] see *Abbreviations* on page 18;
[(b)] duration of each production stage.

3.1.2 Preparation of the simulated systems

Hydrogen atoms were added to all the models: we assumed the standard protonation of the nucleotides at pH 7.4 and the aminoglycosides in a fully protonated state (i.e., all amine groups became NH_3^+).[12] The systems were subsequently neutralized by adding sodium ions (Na^+) in the electrostatic potential minima with the `LEaP` program from Amber 9[143] (or Amber 11[144] in the case of WT_KAN simulation), resulting in 44, 34 and 36 ions added to the bare RNA, complexes with PAR and the complex with KAN, respectively. The neutralized systems were then submerged in water boxes of dimensions $92 \times 69 \times 69$ Å^3 using the TIP3P water model.[145] Finally, random water molecules were substituted for 39 Na^+ and 39 Cl^- ions, to create the ionic strength of ~150 mM, which resembles the cellular environment.

In the crystal structure used for the WT_KAN simulation an additional KAN molecule is bound non-specifically. This was removed from the structure prior to the system setup. For all the RNA systems we have used the Amber ff99[146] force field. Parameters for the antibiotics were prepared in Amber and the parameterization process is described below, in Section 3.3 *Free aminoglycosides*.

3.1.3 Simulation conditions

The simulation protocol was specifically designed for a system containing short RNA helix. Due to complicated RNA architectures and high charge, the procedure has to be different than for proteins.[147]

1. The **minimization** was performed by `sander` program from Amber: 8 000 steps of SD and 2 000 steps of CG. All heavy atoms (i.e., not hydrogens) were restrained with harmonic position constraints (with constraint coefficient $k = 100\ \mathrm{kcal/(mol{\cdot}\AA^2)}$).

2. During the **thermalization**, temperature was gradually increased from 30 to 310 K and the constraints were weakened to $k = 50\ \mathrm{kcal/(mol{\cdot}\AA^2)}$ for the first 85 ps, and $k = 25\ \mathrm{kcal/(mol{\cdot}\AA^2)}$ for the next 35 ps.

3. Next, there was the **two-step equilibration**: during the first 300 ps the constraints were gradually weakened and for the last 600 ps only the terminal nucleotides ($k = 0.35\ \mathrm{kcal/(mol{\cdot}\AA^2)}$) plus the phosphate atoms (P; $k = 0.25\ \mathrm{kcal/(mol{\cdot}\AA^2)}$) were assigned small constraints. The values of these constraints was adjusted so as to match the thermal fluctuations (i.e., the β-factors) for the analogous atoms from the crystal structure of the whole 30S subunit (PDB id: 1IBK[148]).

4. The last stage, **production**, was performed with the same constraints as for the second equilibration step, and lasted 20 or 25 ns (see Table 3.1). In the tests, the removal of constraints on terminal nucleotides resulted in fraying ends, and because in the whole ribosome, the A-site is contained within a longer helix, we did not want to allow for the fraying termini.

The simulations were performed in NAMD[149] in the NPT ensemble: the temperature was controlled by Langevin thermostat (collision frequency of 1 per ps), and the pressure (1 atmosphere) was maintained using the Langevin piston method.[121,150] Electrostatic interactions were calculated using PME,[119] and a cutoff of 12 Å was used to limit the direct space sum. The SHAKE algorithm[123] was used to constrain all the bonds with hydrogen atoms, which enabled a 2-fs long timestep.

3.1.4 Choice of the force field

The simulations of RNA were conducted in ff99 Amber force field. During our study newer force fields parameters became available, namely the ff10[151–153] model, with an improved description of the torsional angles for the sugar–base (χ angle) and sugar–phosphate (α and γ angles) linkages in long simulations (> 50 ns). Our trajectories were, however, shorter (maximum 25 ns), and therefore, the used parameters should not bias the outcome. Banáš *et al.*[152] have shown that simulations of RNA in ff99 could lead to a severe disruption of the secondary helical structure, due to misparameterized χ angle. However, throughout our simulations, we have not observed the described switch of χ values from $\sim -150°$ to $\sim -90°$ (the mean value calculated for all the residues for the whole simulation time reached maximum of $-132°$ for the G1491U simulation). Moreover, recently we have also performed

100 ns-long simulations of the bare RNA A-site model in both ff99 and ff10, and the analysis of the χ angle showed no difference for the first 50 ns (data not shown).

3.1.5 Reliability of the dynamics of the A-site model

Inside the small ribosomal subunit, the A-site interacts with other parts of the ribosome, thus the internal dynamics of the nucleotides observed in MD simulations of the isolated A-site model could in principle differ from their behavior in the whole ribosome. To investigate the structural environment of the bulged bases A1492 and A1493 inside the ribosome, we gathered known ribosomal structures from PDB and calculated the solvent accessible surface areas (SASA) of these bases. Figure 3.2 shows that the majority of the X-ray SASA values are within the range or larger than the values measured in the MD simulation of the free A-site. The only points that are lower than the MD-derived values originate from the ribosomal complexes with the drugs binding in the A-site or in its immediate vicinity, which naturally diminishes the solvent accessibility of A1492 and A1493. Thus, following the previous experimental studies[140,141] and our analysis, we assumed that this short fragment of 16S rRNA constitutes a reasonable model for our studies.

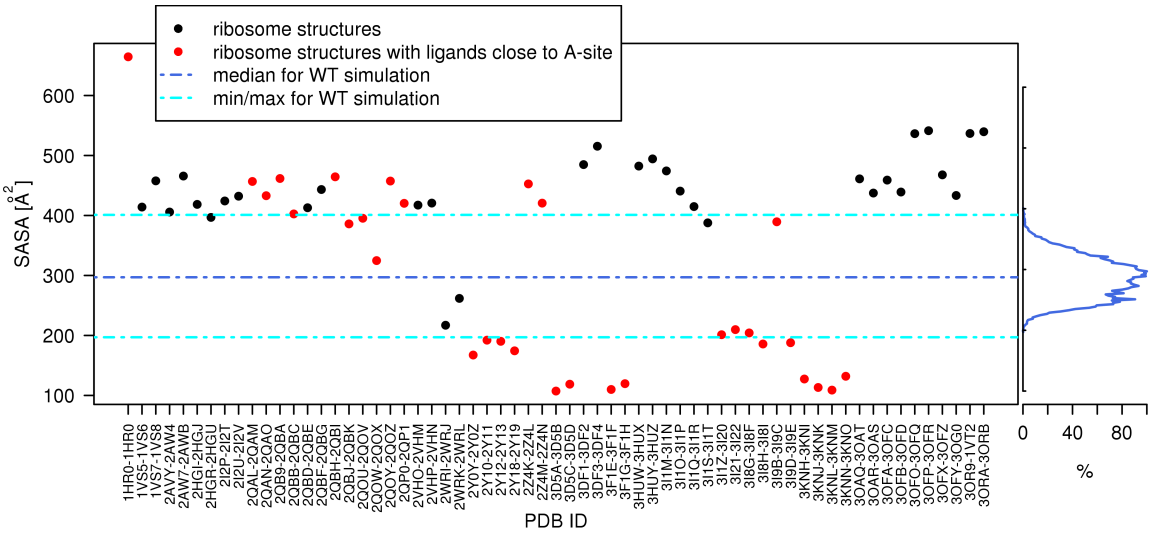


Figure 3.2: **Solvent accessible surface area (SASA [Å^2]) of the adenines A1492 and A1493 in different ribosomal X-ray structures** (identified by their PDB codes on the *x* axis); some ribosomal structures contain ligands bound in the A-site neighborhood, which significantly decreases the SASA values. The right panel shows the histogram of SASA values measured in the simulation of the wild-type (WT) A-site model.

3.2 Aminoglycoside modifying enzymes

3.2.1 Starting structures

From all structures of AMEs available in PDB, we have chosen their complexes with aminoglycosides that had a resolution below or equal to 2.5 Å. For the MD simulations only three of them were taken, each one from a different AME family. Table 3.2 lists all the structures chosen for further studies.

To make the simulations comparable, we had to modify some of the selected structures. In the complex of AAC(6′)-Ib (PDB ID 1V0C), kanamycin C was present instead of kanamycin A. Two substituents were swapped to regain kanamycin A: 6′-OH and 2′-NH_2. Also in this enzyme, there were three amino acids missing: Gly 71, Glu 72, and Glu 73, located in the loop above the aminoglycoside binding region. Initial positions of these residues were taken from another X-ray structure, 2BUE, which contained the same enzyme but complexed with a different aminoglycoside.

The structures of two other enzymes, APH(3′,5″)-IIIa (PDB ID 1L8T) and ANT(4′) (PDB ID 1KNY), contained the appropriate antibiotic, but an inactive form of the cofactor: AMPCPP and ADP, respectively. In the case of ANT, the change into the active cofactor (ATP) involved changing only one atom from C to O. In the case of APH, a phosphate group was added to the ADP molecule.

Table 3.2: **Structures of AMEs that were used in the study.**

PDB ID	enzyme	cofactor	aminoglycoside	ions	reference
1V0C	AAC(6′)-Ib (†*)	ACO	kanamycin C	Ca^{2+}	[60]
2BUE	AAC(6′)-Ib	ACO	ribostamycin	Ca^{2+}	[60]
1L8T	APH(3′,5″)-IIIa (†*)	ADP	kanamycin A	Mg^{2+}	[74]
1KNY	ANT(4′) (†*)	AMPCPP	kanamycin A	Mg^{2+}	[68]
1L8T	APH(3′,5″)-IIIa (†)	ADP	kanamycin A	Mg^{2+}	[74]
1J7L	APH(3′,5″)-IIIa (†)	ADP	—	Mg^{2+}	[74]
1S3Z	AAC(6′)-Iy (†)	coenzyme A	ribostamycin	—	[154]
1M4I	AAC(2′)-Ic (†)	coenzyme A	kanamycin A	—	[60]

(*) initial structures for the simulations
(†) structures analyzed with the use of APBS and Surface Diver

3.2.2 Preparation of the simulated systems

We wanted to perform the simulations under conditions similar to the cellular environment (pH 7.4), and therefore, kanamycin A was in a fully protonated state (i.e., with all amine groups being NH_3^+).[12,155] For proteins, adding hydrogen atoms was not straightforward. Initially, we used the PDB2PQR server[156] together with the PropKa software[157] (version 1.0) to determine the protonation states of every amino acid in the unliganded enzymes. Employing the same approach for the structures complexed with kanamycin A resulted in protonation states of the residues of the enzyme that disfavored the formation of important hydrogen bonds with the ligand. Inside the binding sites of AMEs, many Glu and Asp residues create an environment that can be too complicated for automated predictors, such as PropKa. These specific protonation states in such environments have been studied[12,67,158,159] and it has been shown that this is connected with partial protonation of an aminoglycoside during its binding to an enzyme. Hence, we have decided to manually

Table 3.3: **MD simulations of representative AMEs.**

system description[a]		simulation name	length [ns]
AAC + ACO	+ KAN	AAC_HOLO	3 x 20
	− KAN	AAC_APO	3 x 20
APH + ATP	+ KAN	APH_HOLO	3 x 20
	− KAN	APH_APO	3 x 20
ANT + ATP	+ KAN	ANT_1KAN	3 x 20
	+ 2 x KAN	ANT_HOLO	3 x 20
	− KAN	ANT_APO	3 x 20

[a] Molecules are described by abbreviations (see Section *Abbreviations* on page 18). See Table 3.2 for details of the structures.

place hydrogen atoms of the residues inside the binding sites, based on experimental data and visualization of the structures. As a result, all of the Glus and Asps were negatively charged, like in water environment, apart from Asp 50 (named ASH) in both monomers of the ANT dimer, which was assigned a neutral state.

Each of the initial structures was neutralized by addition of sodium ions, Na^+, and submerged in a truncated octahedron of TIP3P water molecules. The minimal distance between the solute and the edge of the system was 15 Å.

3.2.3 Simulation conditions

We used the newest available Amber force field dedicated for proteins, ff03.r1.[160,161] The simulations were performed with the use of `pmemd` (Particle Mesh Ewald Molecular Dynamics) and `sander` programs from Amber11 suite. The simulation protocol was the following:

1. The **minimization** consisted of two parts: *(i)* 3000 steps of SD search and 2000 steps of CG algorithm with harmonic restraints on the solute (force constant of 2.0 kcal/(mol Å^2)); *(ii)* afterwards, no constraints were applied, the whole system underwent 7000 steps of SD search, followed by 3000 steps of CG.

2. The minimized structures were then subjected to slow **heating** to 300 K during 50 ps of MD with restraints on the solute (force constant of 2.0 kcal/(mol Å^2)).

3. The following **equilibration** stage was also two-fold: *(i)* during the first 50 ps constraints of 2.0 kcal/(mol Å^2) were applied to the solute; *(ii)* for the remaining 1 ns the constraints were four times smaller. Equilibration was performed in the NPT ensemble, with Langevin thermostat and isotropic position scaling for the constant pressure.

4. The main part, **production** stage was performed in the microcanonical ensemble (NVE). A timestep of 1 fs was used together with a lower tolerance of the SHAKE algorithm (10^{-6} Å, restraining water hydrogen atoms only), in order to ensure a stable temperature. The electrostatic calculations were performed with the PME approach with a one order lower tolerance of the direct sum (10^{-6}), which decreases RMSD of force

errors.[144] Both, the temperature and the total potential energy were stable, with fluctuations around the mean of less than 0.5% and 0.01%, respectively.

For each of the systems, the simulation was repeated three times, each time with different initial velocities. This increases the sampling of the molecular conformational space and the statistical relevance of the findings. Table 3.3 presents a summary of all the simulations.

3.3 Free aminoglycosides

3.3.1 Simulation setup

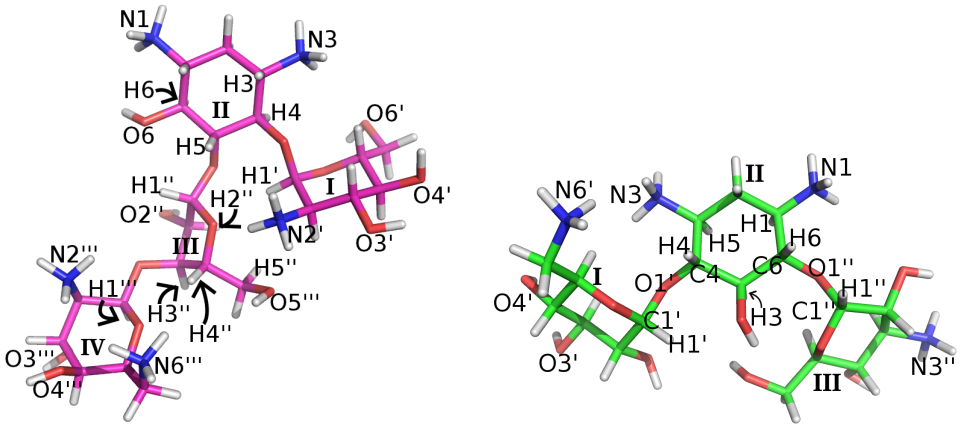
Simulations of the two aminoglycosides, paromomycin (4,5-disubstituted DOS) and kanamycin A (4,6-disubstituted DOS), were performed in explicit water, in order to check the correctness of the parameterization. The initial positions of the atoms of the drugs were taken from the structures of their complexes that had the best resolution: 1J7T[142] for PAR (2.5 Å), and 1ND4[162] (2.1 Å) for KAN. Almost all of the amine groups of paromomycin[12,163] and kanamycin A[155] have pKa values > 7.4, which means that these are protonated at physiological pH. The pKa value of one substituent, namely group 3-NH_2 in both of these antibiotics (see Figure 1.1), was found to be slightly lower than 7 in water, however, it was shown that the environment of the RNA A-site or enzymes promotes full protonation of aminoglycosides.[12] Thus, for the simulations of the drugs alone, we have added hydrogens to all of their amine groups, in order to use these structures and parameters in the simulations of the complexes. This resulted in the total charge of $+5$ and $+4$ for PAR and KAN, respectively.

The parameters were based on GAFF (General Amber Force Field).[164] For paromomycin, the partial atomic charges were taken from a standard set called AM1BCC.[165] In the case of kanamycin A, charges were fitted by `antechamber` with the RESP algorithm, based on the electrostatic potential calculated by Gaussian[166] with the GTO 6–31G* basis set.

Both drugs were neutralized by adding chloride ions, Cl^-, prior to putting them into a truncated octahedron of TIP3P water molecules. The minimal distance between the drug and water edge was 15 Å. The simulations of aminoglycosides were carried out with the use of `sander` from Amber11. The protocol was essentially the same as for the simulations of the enzymes. We performed three simulations per aminoglycoside, each one with different initial velocities, each production lasting 10 ns. The length of each simulation was short because our aim was only to validate the force field parameters and not to thoroughly investigate the conformational phase space of these molecules.

3.3.2 Parameterization validation

Aminoglycoside antibiotics posses several rings that are connected through flexible linkages (see Figure 3.3). The different conformations they adopt can be quantified by measuring the dihedral angles of the atoms in the linkages, as proposed in Refs.[52,167] The general definitions are: ϕ($H1_{sugar}$–$C1_{sugar}$–$O1_{sugar}$–Cx) and ψ($C1_{sugar}$–$O1_{sugar}$–Cx–Hx), where "x" depends on the linkage. For example, for linkage I/II in kanamycin A: ϕ(H1′–C1′–O1′–C4) and ψ(C1′–O1′–C4–H4); see Figure 3.3B for the numbering of atoms. Figure 3.4 shows the results for the two studied aminoglycosides; the measurements were done for MD simulations in explicit water. Each linkage has one conformation that is the most populated, but the range of the observed dihedral angles is quite wide. Importantly, linkage I/II behaves similarly for both studied antibiotics. This is expected, since rings I and II are common to the majority of aminoglycosides. These two drugs, however, belong to two different subgroups — kanamycin A is a 4,6-DOS, while paromomycin is a 4,5-DOS. Therefore, linkages II/III scan different parts of the ϕ/ψ conformational space.



(A) paromomycin (B) kanamycin A

Figure 3.3: **Starting structures of the two aminoglycoside representatives**, shown as sticks, colored by atom type: C — magenta or green, N — blue, O — red and H — white. The rings are numbered with Arabic numbers and the atoms referred to in the text are labeled.

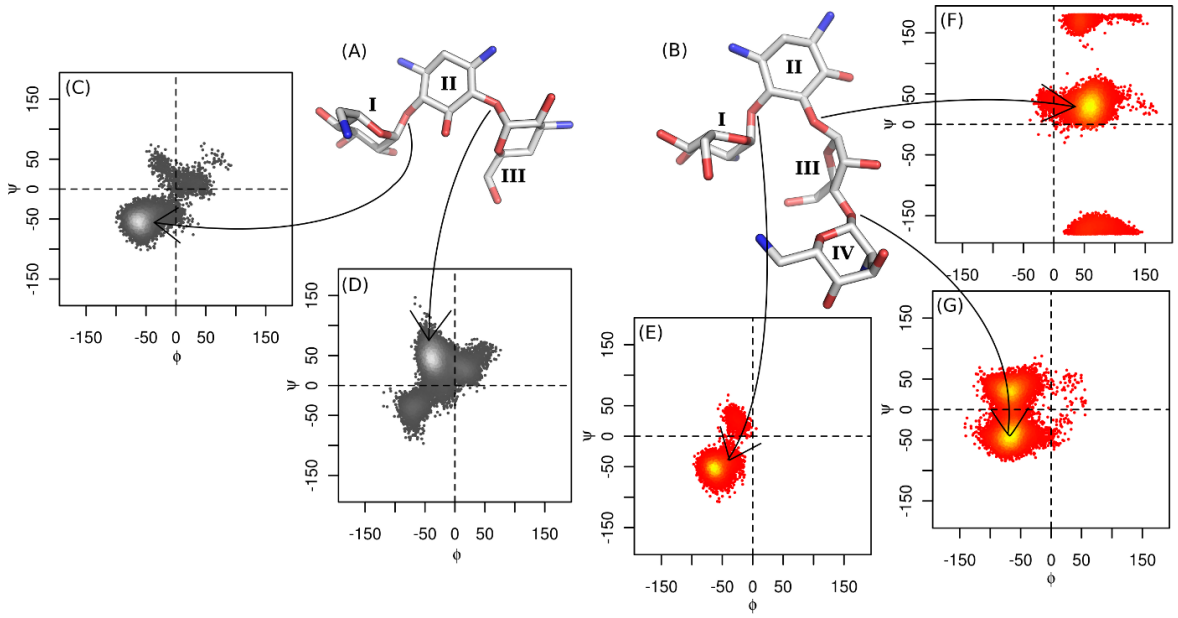


Figure 3.4: **Conformations of the studied aminoglycosides**, *(A)* kanamycin A and *(B)* paromomycin (only the representative conformations are shown), measured by dihedral angles of linkages: *(C),(E)* linkage I/II, *(D),(F)* linkage II/III, and *(G)* linkage III/IV.

There were no experimental data exactly for paromomycin, thus we have compared the MD-derived values with the NMR data for neomycin B,[167] which differs from paromomycin only in one substituent (6′-NH_2 instead of 6′-OH). Moreover, these experiments were conducted at pH values of 4.7 and 9.7, while in the simulations, we mimicked pH 7.4. The NMR experiments[52,167] showed that linkage I/II for both aminoglycosides acquired most often a syn-ψ orientation ($\psi \approx 0°$), and the population of anti-ψ orientation ($\psi \approx 180°$ or $\psi \approx -180°$) had a different pH dependence for these two drugs. In the case of kanamycin A, the anti-ψ orientation was sampled most often at low pH, and it diminished at high pH. On the contrary, for neomycin B, only at high pH the anti-ψ orientation of linkage I/II was measurable. In the simulations, we observed only the syn-ψ orientation (Figure 3.4C,E).

Linkage II/III was found to acquire only syn-ψ orientation for kanamycin A,[52] while

for neomycin B the anti-ψ orientation was sampled for 17% of the cases.[167] Moreover, in experiments, the neomycin B linkage II/III occupied a non-*exo* conformation (i.e., ϕ value of $\sim -10°$) for 35% of the time. We observed the same scheme for kanamycin A and paromomycin in the MD simulations; however, the occupancy of the non-*exo* conformation in par was lower (only $\sim$ 6%; Figure 3.4D,F). Finally, the measured ψ angle for linkage III/IV of neomycin B[167] was close to $50°$ for half of the sampled conformations, and close to $-50°$ for the other half. In the simulations, paromomycin linkage III/IV sampled the conformations with $\psi \in [-30; -70]°$ more often then the conformations with $\psi \in [30; 70]°$ (45% vs. 23%) (Figure 3.4G).

In addition, we measured the distances between the selected hydrogen atoms and compared them with the experimental data.[52,167] Figures 3.5 and 3.6 present the experimental and MD-derived distances for paromomycin and kanamycin A, respectively. In the case of paromomycin, we found that the distances 7 and 10 are larger in our simulations than in the experiment. These refer to the relative orientation of rings II and III (Figure 3.4F), where the occupancy of the non-*exo* conformation in the simulation was much higher in the experiment than in our simulations, probably due to limited sampling. Moreover, for a similar reason distance 13 was larger in the simulation than in the experiment. In general, however, the average MD distances were in sufficient agreement with the NMR experiments.

In the case of kanamycin A, the agreement between the experiment and the simulations was good, except for the distances 2 and 6, which were larger in the simulations. This is most probably caused by no anti-ψ conformations sampled in the trajectories (Figure 3.4C–D), in which these distances are shorter.

Overall, for the purpose of our simulations of aminoglycosides in the complexes with the RNA A-site and AMEs, the force field parameters seem good enough. The starting conformations of aminoglycosides were taken from the X-ray structures of the complexes, and the short simulations were conducted only to ensure that the parameterization was reliable, not to study the conformational transitions of these drugs.

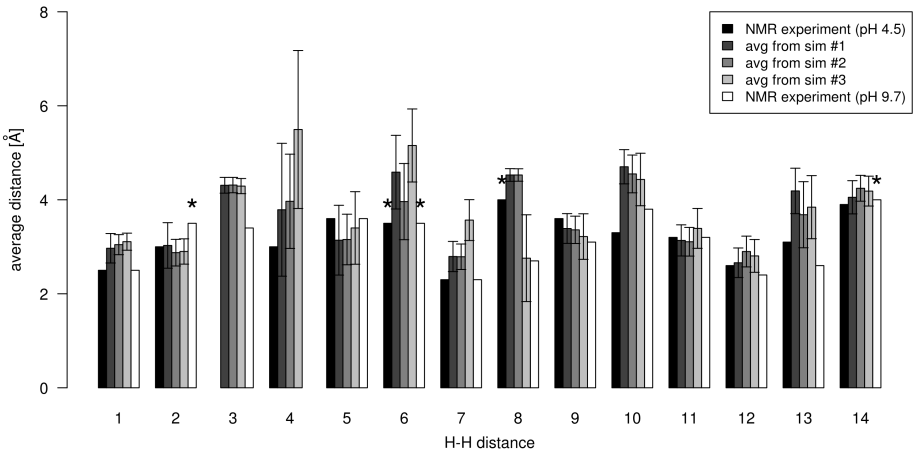


Figure 3.5: **Distances between chosen hydrogen atoms of paromomycin,** averaged over the trajectories, compared with the values from an NMR study of neomycin B, which differs from paromomycin only with one chemical group (6-OH instead of 6-NH_2). The experimental values are taken from Table 1 in Ref.[167] and are numbered as follows: 1 stands for the distance between H1′ and H4, 2 — H1′–H5, 3 — H1′–H3, 4 — H1′–H5″, 5 — H1′–H2″, 6 — H1′–H3″, 7 — H1″–H5, 8 — H1″–H4, 9 — H1″–H6, 10 — H2″–H6, 11 — H1″–H4″, 12 — H1‴–H3″, 13 — H1‴–H2″, 14 — H1‴–H4″. See Figure 3.3A for atom labeling. The asterisks (*) mark the distances for which only lower bounds were given in the experimental study.[167]

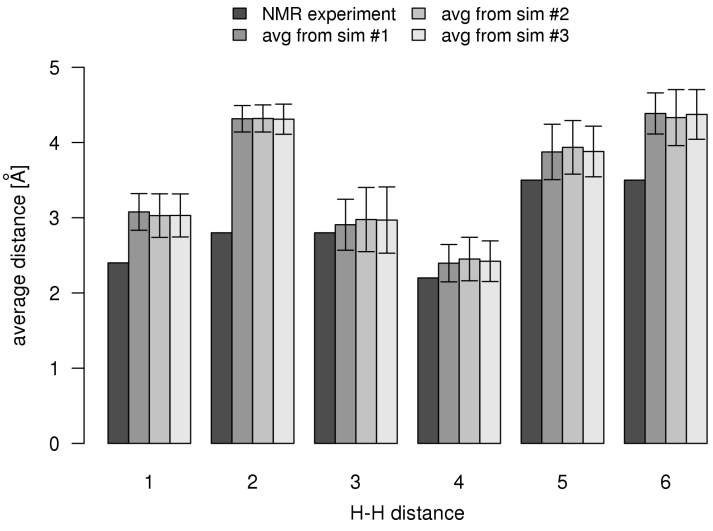


Figure 3.6: **Distances between chosen hydrogen atoms of kanamycin A** — a comparison of values obtained from the simulations and NMR experiments at pH 7.7, as shown in Figure S2 in Ref.[52] The H-H distances are numbered as follows: 1 stands for the distance between H1′ and H4; 2 — H1′–H3; 3 — H1′–H5; 4 — H1″–H6; 5 — H1″–H5; 6 — H1″–H1. See Figure 3.3B for atom labeling.

3.4 Data analysis

For visualization of the trajectories, we used VMD[168] and PyMol Molecular Graphics System (Schrödinger, LLC., www.pymol.org). The plots and graphs were produced in the R environment.[169] VMD was also used to measure the following quantities (through custom Tcl/Tk scripts):

- the dihedral angles;
- the average density of water or ions inside the binding sites;
- SASA;
- and the stacking interactions.

The basic measures, i.e., RMSD and RMSF, were calculated with either `g_rms` and `g_rmsf` programs of the GROMACS package[170–172] or `ptraj` of AmberTools. `ptraj` was also used for the following analyses:

- *clustering of conformations*: average linkage method was used; we performed this hierarchical clustering till all the clusters were merged and then we chose the number of clusters that would give the optimal distribution of the conformations, based on the methodology described in Ref.;[173]
- *finding hydrogen bonds* between the ligand and the receptor; maximal distance between the donor and the acceptor was set to 3.5 Å and the angle was minimum 120°;
- *measuring distances* between atoms or centers of mass of residues;
- *outputting the average conformations*.

To analyze the tertiary structures of the simulated RNA, we used the 3DNA software,[174] which measures various helical parameters of nucleic acids, including the glycosidic angle χ, the number of hydrogen bonds between bases in a pair or the opening angle of a base pair. The method is based on a geometrical analysis and comparison of the nucleotide conformation from the simulation to a template structure.

From the pharmacological point of view, the volume of a binding pocket is an important feature of a receptor because it can be critical when designing drugs and checking their efficiency. It is easy to find and define a binding pocket on a surface of a protein (e.g., Ref.[175]), but it is not straightforward for RNA structures since there are many cavities on a surface of e.g., a double helix. We have chosen the POVME software[176] for calculating the volume of binding sites because it offers a generic approach and the area of the binding pocket can be precisely defined by the user. The implemented algorithm measures the volume of a chosen cavity by placing beads on a grid wherever the bead does not contact any atom of the molecule. Thus, the sum of placed beads times the grid resolution yields the total volume. The volumes were measured without the aminoglycoside inside the binding site.

PCA was done with the use of `bio3d` package[177] for R. Adaptive Poisson-Boltzmann Solver (APBS)[178] was used to calculate the ESP of the X-ray structures of AME ternary complexes (enzyme:cofactor:aminoglycoside). However, for the calculations, we deleted the aminoglycoside that was in the X-ray structure. Next, the Surface Diver software[179] was used to compare these potentials in relation to an aminoglycoside core (i.e., neamine part) inside the binding cleft. Surface Diver performs a decomposition of ESP into spherical harmonics, and therefore, the compared molecules do not need to be structurally or sequence related.

The binding free energy and its per-residue decomposition was computed by MMPBSA.py, a part of AmberTools, on total of 3333 frames from each trajectory, i.e., every 6 ps. For calculating the electrostatic interactions, we chose the simpler generalized Born model[132] instead of the Poisson-Boltzmann model[180] because of the large sizes of simulated systems and large

number of simulations. The calculated energy accounts only for the enthalpic part of the total binding free energy ($E_{\mathrm{calc}} \equiv H$; see Chapter 2, Section 2.8 *Assessment of the binding free energy*), and therefore we used the calculated values only to compare the contributions coming from single residues. Moreover, the experimental data on the free energy of binding of aminoglycosides to AMEs is scarce, thus we could not relate to them the calculated energy values.

In the AAC_HOLO and APH_HOLO simulations, some water molecules close to kanamycin A atoms were especially stable (calculated β–factors were less than 30 Å^2), thus we decided to include them as part of the receptor when calculating the energies. However, in each of the three separate APH_HOLO simulations, these were different water molecules. Therefore, we performed three separate calculations for each APH_HOLO simulation and then we combined the results using the weighted mean (μ^*) and the unbiased error estimation (σ):

$$\mu^* = \frac{\Sigma_{i=1}^{N} w_i x_i}{V_1} ,$$

$$\sigma^2 = \frac{V_1}{V_1^2 - V_2} \sum_{i=1}^{N} w_i (x_i - \mu^*)^2 ,$$

where w_i is the weight of element x_i (here the weights are reciprocals of standard errors of energy values calculated for each of the simulations); $N = 3$ is the number of simulations; $V_1 = \Sigma_i w_i$ and $V_2 = \Sigma_i w_i^2$.

To determine which residues are contributing the most to kanamycin A binding to AMEs, we used a measure introduced in Ref.[181] The amino acid efficiency is defined as follows:

$$EFF = \frac{\Delta E_{res}}{N_{hvy}} ,$$

where ΔE_{res} is a total contribution of a residue (difference of the MMGBSA energy for the residue in the APO and HOLO states), and N_{hvy} is a number of heavy atoms in this residue.

Chapter 4

Results and discussion

We used several tools of theoretical biophysics to investigate the binding sites of aminoglycosides located in different biomolecules: the ribosomal RNA A-site and three AME representatives. We focused on two types of bacterial resistance: *(i)* mutations in the RNA binding site; and *(ii)* modification of aminoglycosides by the bacterial enzymes. We analyzed the "static" X-ray structures but the most important conclusions are based on the results from molecular dynamics simulations of these different molecules and their complexes with two aminoglycoside representatives, kanamycin A and paromomycin. This wide range of structures and the biochemical differences between them, required the usage of various methods. We also developed a new method for conformational analysis. In this study, we aimed at characterizing and comparing the environments that these biomolecules provide and the types of interactions formed with aminoglycosides.

This chapter is organized as follows. First, we present our new tool for analyzing internal dynamics of biomolecules, and we demonstrate its usability through a series of tests in Section 4.1 *New method for identifying similarly moving atoms in molecular conformation sets*. Next, in Section 4.2 *Bacterial resistance mechanism 1: ribosomal RNA A-site mutations*, we describe and discuss the results from the MD simulations of the mutated A-site models. Finally, Section 4.3 *Bacterial resistance mechanism 2: enzymatic aminoglycoside modification* focuses on the MD simulations of AMEs, where the results are explained and discussed.

4.1 New method for identifying similarly moving atoms in molecular conformation sets

Currently, the simulations produce even hundreds of thousands of molecular conformations that have to be carefully analyzed. Especially in classical MD, these conformations are most of the time very similar to one another, due to the intrinsic limitation of sampling restricted to only a couple of local energetical minima. Therefore, noticing a larger transition can be difficult. To make it easier, one often decreases the dimensionality of the description of molecular motion. One of the methods is searching for similarity in atomic motions. The classical approach involves calculation of cross-correlation coefficients, which yield DCCM. This approach, however, has some limitations, as explained in Chapter 2, Section 2.5.3 *Correlated motions*. Therefore, we proposed an alternative method, where the molecule is divided into so-called "dynamic domains", based on its conformations. These dynamic domains can be defined as parts of the molecule that stay internally rigid but move in relation to one another. The idea is not new — the first algorithm and software, Hingefind, was proposed by W. Wriggers and K. Schulten in 1997.[182] Their algorithm compares two conformations of a protein by least-square fitting, and then finds the stable fragments and defines a rotation axis. DomainFinder, designed by Konrad Hinsen,[183,184] is perhaps more commonly used. Based on a normal mode analysis, DomainFinder finds the directions of harmonic motions of atoms around their energetic minimum.[136] find.core from the R package *bio3d*,[177] finds the atoms that are invariant in a molecule, based on its conformations or alignment to a similar set of structures. FindCore algorithm[185] analyzes variation of distances between pairs of atoms and also outputs the "core atoms", i.e., the atoms that change their positions the least. It does not, however, predict a clear division into dynamic domains. The CYRANGE program[186] gives this division explicitly, but it was optimized to analyze the data from NMR experiments, not MD simulations. Next, the DynDom database[187,188] calculates and stores a division into dynamic domains for proteins that have two experimentally resolved conformations. A more recent server, PiSQRD,[189,190] uses a covariance matrix to extract low-energy modes of motions, and, based on these modes, the atoms are grouped into dynamic domains.

The presented list of software and methods shows that a lot has been already done to analyze the internal dynamics of molecules. However, none of these approaches was designed to deal with large datasets that come from simulations — they focus on experimentally resolved conformations. Moreover, all of the mentioned software analyzes only proteins. Therefore, we found a need to create a similar tool that would deal well with data produced by simulations and that could be also applied for nucleic acids. Additionally, we wanted to improve the way of detecting the correlations of motions and find a method that does not depend on the order of conformations and the type of molecules being analyzed.

Below, we describe our new algorithm, the implementation (Geometrically Stable Substructures, GeoStaS) and its testing on experimental data (obtained from NMR) and MD simulations, of proteins, nucleic acids and their complexes. The software is open source and freely available for download: `http://bitbucket.org/jrom/geostas` or `http://bionano.icm.edu.pl/Software/GeoStaS`.

4.1.1 GeoStaS algorithm

The workflow of GeoStaS algorithm is presented in Figure 4.1. The algorithm is based on finding similar atomic motions. Instead of concentrating on the whole molecule, we extract trajectories of individual atoms from the provided set of molecular conformations (which can be a trajectory from a simulation or an ensemble of experimentally resolved

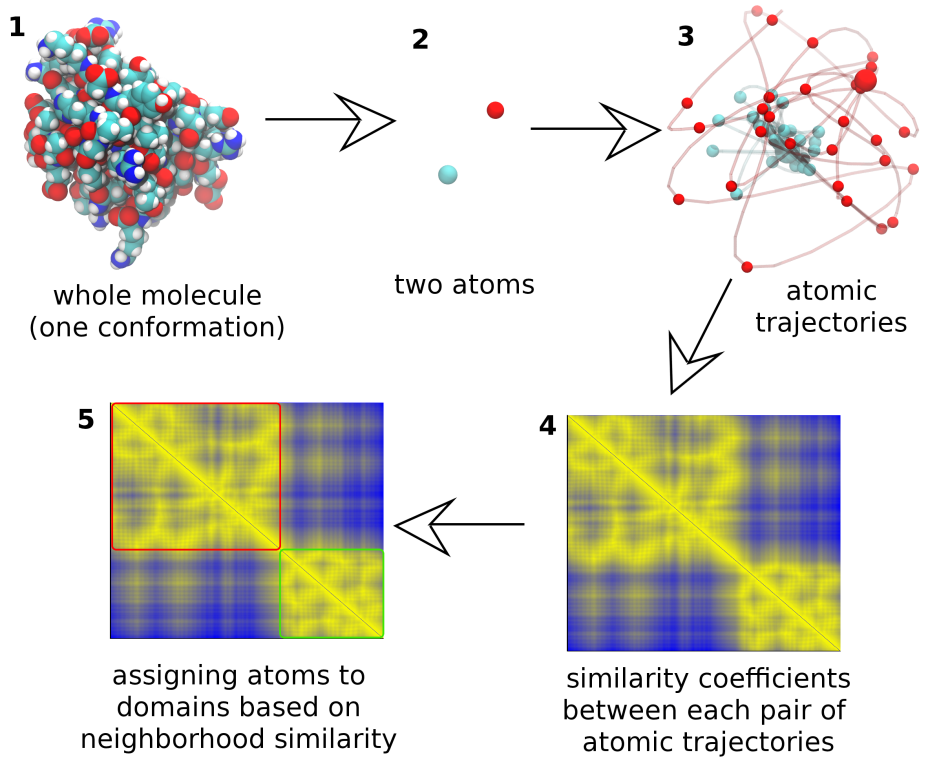


Figure 4.1: **Schematic representation of the workflow in the GeoStaS algorithm.** Detailed description is given in the main text.

structures). These traces of atomic movements are then compared strictly geometrically, through finding an isometry, i.e., translation plus rotation, that would minimize the distance between them. To find such translation, the conformations are first repositioned with regard to coordinates of the first conformation. Finding the best rotation is more complicated. In order to simplify the calculations, this is done in the quaternion space.

Rotation in quaternion space

Quaternions are an extension to complex numbers, introduced by sir William Rowan Hamilton in the 19$^{\text{th}}$ century.[191] A quaternion is defined as a four tuple:

$$q = q_0 + \vec{q} = q_0 + q_1 \cdot \vec{i} + q_2 \cdot \vec{j} + q_3 \cdot \vec{k} \,, \tag{4.1}$$

where q_0 is named the scalar part and $\vec{q}$ is the vector part. Multiplication of quaternions is non-commutative and it is defined by two rules:

$$\vec{i}^2 = \vec{j}^2 = \vec{k}^2 = -1 \,,$$
$$\vec{i}\,\vec{j}\,\vec{k} = -1 \,.$$

Then, multiplying two quaternions, p and q, gives:

$$(p_0 + \vec{p})(q_0 + \vec{q}) = p_0 q_0 - \vec{p} \cdot \vec{q} + p_0 \vec{q} + q_0 \vec{p} + \vec{p} \times \vec{q} \,.$$

The conjugate of a q is $q^* = q_0 - \vec{q}$, and their multiplication:

$$q q^* = q_0^2 + q_1^2 + q_2^2 + q_3^2$$

gives a nonnegative real number.[192]

In the case of GeoStaS algorithm, the key feature of quaternions is that when constructing a quaternionic operator and applying it to a vector $\vec{v}$ in $\mathbb{R}^3$, we obtain the rotated vector $\vec{v}\,'$ by angle α:

$$\vec{v}\,' = q\ \vec{v}\ q^{-1}\ , \tag{4.2}$$

$$q = cos(\alpha/2) + \vec{a} \cdot sin(\alpha/2)\ , \tag{4.3}$$

where $\vec{a}$ is the direction of $\vec{q}$, i.e., the normalized vector part of q; and $q^{-1} = q^*/|q|^2$.

The standard approach for performing a rotation in Cartesian space uses a rotation matrix 3×3 that contains trigonometric functions of rotation angle, which are difficult to implement. In the quaternion space, one needs only four parameters (one quaternion) and there are no trigonometric functions. Kneller *et al.*[193] have presented an elegant way of implementing this search for the best rotation in quaternion space. It can be reduced to a simple eigenproblem:

$$\hat{M} \cdot q = \lambda\ q\ , \tag{4.4}$$

where q is the quaternion that describes the rotation; and λ is a diagonal matrix. Matrix $\hat{M}$ is of the following form:[193]

$$\hat{M} = \sum_{k=1}^{N} \begin{bmatrix} \left(\vec{x}_i(k) - \vec{x}_j(k)\right)^2 & \vec{u}_{ij}(k)^T \\ \vec{u}_{ij}(k) & \hat{P}_{ij}(k) \end{bmatrix}\ ,$$

$$\vec{u}_{ij}(k) = \vec{x}_i(k) \times \vec{x}_j(k)\ ,$$

$$\hat{P}_{ij}(k) = \vec{x}_i(k) \cdot \vec{x}_j(k)^T + \vec{x}_i(k)^T \cdot \vec{x}_j(k)\ ,$$

where $\vec{x}_i(k)$ and $\vec{x}_j(k)$ describe the positions of the two atoms, i and j, respectively, in the conformation k; and ^{T} denotes the transpose of a matrix. For each pair of atoms, their traces are compared, yielding a separate eigenproblem of the form as in Equation 4.4.

The result is a set of four eigenvalues, where the largest value, λ_1, points to the best rotation, while the smallest value, λ_4, shows the worst one. Kneller *et al.* have shown that with the smallest eigenvalue one gets also the normalized distance between the two compared objects:[193]

$$\Delta_\Omega = \left(\frac{M_{11}}{\lambda_4}\right)^{\frac{1}{2}}\ . \tag{4.5}$$

The distance $\Delta_\Omega \in [0;1]$, which discriminates between the parallel ($\Delta_\Omega \sim 0$) or anti-parallel order ($\Delta_\Omega \sim 1$). Since in our algorithm, we want to find the atoms that move in the same direction at the same time (i.e., that are correlated and not *anti*-correlated), we define a similarity coefficient as: $SC = 1 - \Delta_\Omega$, so that higher values would describe better correlations.

Clustering of similarity matrix

These similarity coefficients are calculated per each pair of atomic trajectories and a so-called atomic movement similarity matrix (AMSM) is constructed (see Figure 4.1). Next, this matrix has to be clustered into groups with maximal internal similarity. For this purpose, a modified nearest-neighbor graph algorithm[194] is used. The atoms are represented as nodes of the graph. Edges, weighted with similarity coefficients ($w_{ij} \equiv SC_{ij}$), connect each two atoms that are farther than four residues apart at a sequence level. The similarity coefficients for atoms closer than this limit are naturally very close to 1, thus including these values would bias the outcome of clustering and would not provide significant results. Atoms

(nodes) connected by an edge are called *neighbors*. If additionally the weight w_{ij} of this edge is larger than a given $w_{min} = w_{max} - \Delta w$ (we call Δw the threshold), these atoms are called the *nearest neighbors*.

Scheme of the nearest-neighbor algorithm:

1. within not yet assigned edges find the edge V_{max} with maximal weight, w_{max};
2. take the nodes, a and b, that are connected by V_{max} and create a new cluster C;
3. expand the cluster:
   (a) from not yet assigned neighbors of a or b take each node that has minimum n_{comm} common nearest neighbors with any $g_c \in C$;
   (b) stop when no more neighbors of a or b pass this criterion or there are no more nodes left;
4. repeat from 1 until all nodes are assigned to clusters.

After a series of tests, the value of n_{comm} was set to 80% of the number of the nearest neighbors of $g_c \in C$. The threshold, Δw, is a parameter, adjustable by the user, since it defines the "coarseness" of clustering, and the user has to decide on a more or less detailed division.

Apart from the nearest neighbor algorithm, we also tested a simple hierarchical merging approach. This clustering is performed on the columns of AMSM, not on separate values of the similarity coefficients. In the beginning, each column represents one cluster. In every step of the hierarchical clustering, the two clusters that have the minimal distance are merged. The distance is calculated as a vector difference between the average vectors (columns) of each cluster. However, this type of clustering requires a number of domains to be given as an input. There are some methods to help decide when the clustering is optimal (see e.g., Ref.[173]), and we plan to implement one of them in the nearest future. Although the hierarchical merging clustering algorithm seems promising, we have not fully tested it, and therefore we present only the results obtained with the use of the nearest neighbor clustering.

Automatic mode

The GeoStaS has an automated mode that suggests an optimal division into dynamic domains. In this mode, an output is generated for a set of threshold values (Δw: 0.3, 0.25, 0.2, 0.15 and 0.1) and the associated divisions of the biomolecule are compared. The comparison is based on the RMSD of each of the identified domains:

$$R_{\text{division}} = \frac{1}{N_{res}} \sum_{D} \left\langle \sqrt{\frac{1}{N} \cdot \sum_{i} \left| \vec{x}_{i(k)} - \vec{x}_{i(k=0)} \right|^2} \right\rangle_K , \tag{4.6}$$

where N_{res} and N are the numbers of all residues and all atoms in the molecule, respectively; $D = \{d_j\}$ are the domains that have more than two residues; i traverses all the atoms in the domain d_j; $\vec{x}_{i(k)}$ describes the position of atom i in the conformation k; and $\langle\ \rangle_K$ denotes an average over the set of conformations $K = \{k\}$. The minimal value of R_{division} gives the optimal solution. This can be then used as a starting point for a more detailed analysis.

Implementation

We chose the Java programming language, because it enables easy implementation of the graphical interface, easy development, and it is independent of the operating system.

GeoStaS processes either the text file (in the PDB format — version 3.3* or the binary file (in the DCD format, used e.g., by the popular MD softwares NAMD[149] or CHARMM[195]). The conformations are superimposed onto the first one, in order to remove the rotation and translation of the molecule as a whole. The user has to choose between the automatic and manual mode, and whether to consider only the Cα atoms (or phosphorus in case of nucleic acids) or three atoms per residue (i.e., Cα-C-N for proteins and P-C3′-C4′ for nucleic acids). There are two types of output: the raw AMSM and the PDB file with the molecule divided into dynamic domains.

4.1.2 Tests on experimental and computational data

The tests were performed on more than 30 conformation sets from NMR experiments and on several trajectories from MD simulations. The results showed that GeoStaS is capable of correctly identifying dynamic domains in different structures: proteins, nucleic acids and their complexes. This is a unique feature of GeoStaS, since all of the already available software accepted only protein structures as an input. Moreover, none of the other programs is able to efficiently handle large data files such as trajectories from molecular simulations.

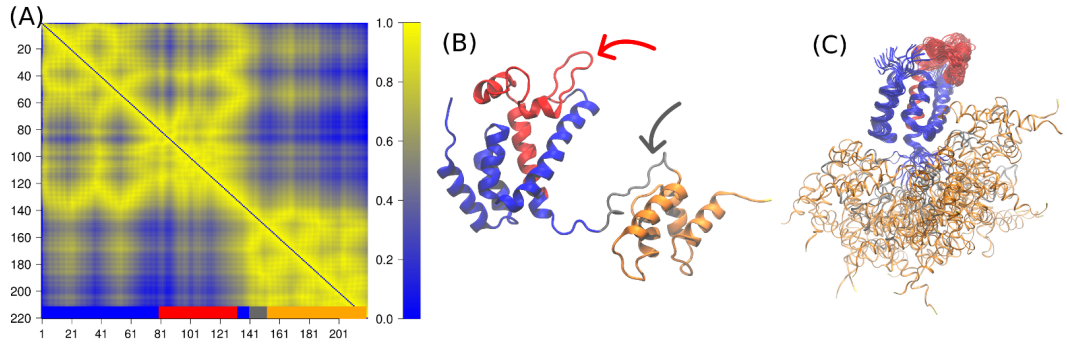


Figure 4.2: **Visualization of the results obtained from GeoStaS, for the NMR ensemble 1D1D**: *(A)* the atomic movement similarity matrix (AMSM), the color bars in the bottom depict which residues are assigned to different domains; *(B)* one molecular conformation colored according to the dynamic domains found by GeoStaS, *(C)* the whole NMR ensemble superimposed on the blue domain.

Experimental data

Exemplary results are presented in Figure 4.2: AMSM, and the molecule colored according to the predicted dynamic domains. The presented molecule is a *Rous sarcoma* virus capsid protein (PDB id 1D1D).[196] Based on the 20 NMR conformations of this molecule, GeoStaS found four domains. This agrees with the analysis of the NOE signals,[196] which identifies the N-terminal (red and blue fragments in Figure 4.2B) and C-terminal (orange fragment) domains as moving independently. Moreover, according to the experiment, one linker between the helices (indicated by the red arrow in Figure 4.2B) is similarly flexible as the linker between the domains (gray arrow). Thus, the movements of the helices connected with this red-colored linker could differ from the movement of the rest of the N-terminal fragment.

The outcome of the analysis of the experimental data was compared with the most recent program, CYRANGE.[186] Figure 4.3 presents the domains found by GeoStaS with the

*www.wwpdb.org/documentation/format33/v3.3.html

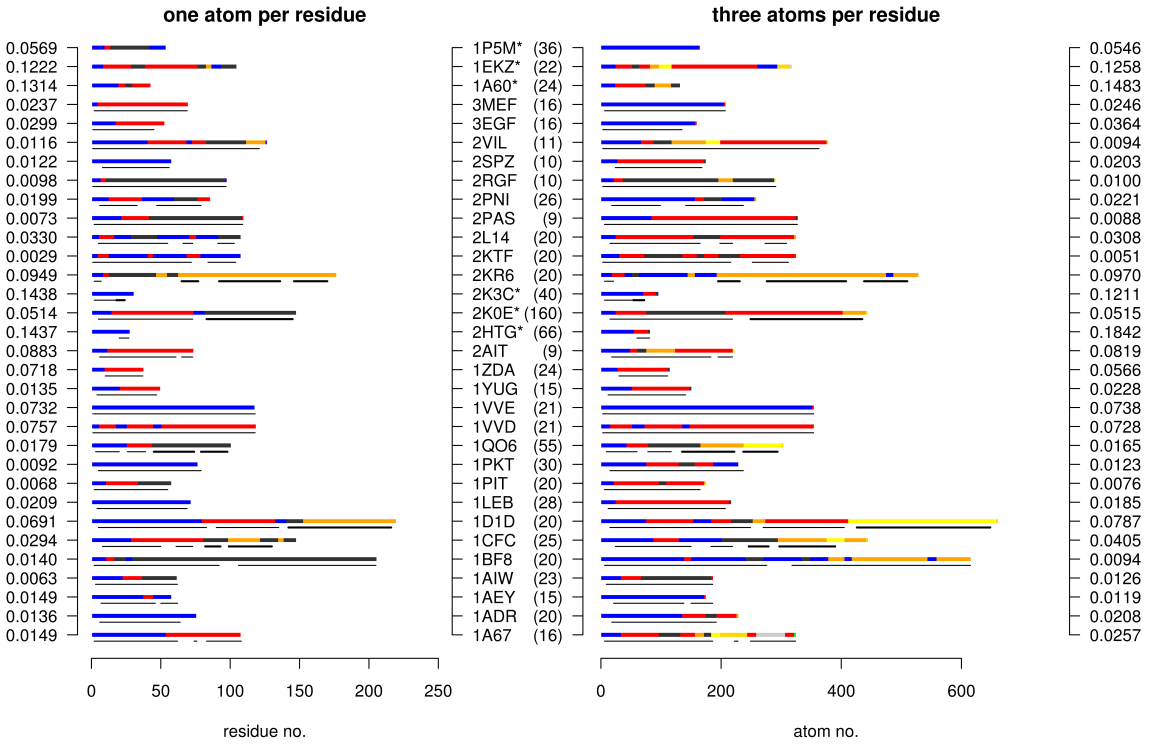


Figure 4.3: **The optimal divisions for all tested NMR ensembles** for *(left)* one– (Cα or P) or *(right)* three– (Cα-C-N or P-C3′-C4′) atoms-per-residue representation; each dynamic domain is colored differently. The NMR ensembles are identified by their PDB codes with the number of conformations given in brackets. The numbers on the far right and left sides show the minimal R_{division} for each of the ensembles. For comparison, the divisions obtained with the CYRANGE program are displayed as black lines below every color bar. CYRANGE gives a range of residues that belong to each domain, and these can have gaps; here, different domains, if identified, are marked with different line widths; the lines for the three-atoms-per-residue mode are scaled by a factor of three, to match the scale of the colored bars, since CYRANGE has only the Cα mode. The structures that were not tested in Ref.[186] are marked with an asterisk. The structures containing nucleic acids, which CYRANGE does not process, could not be compared; these are: 1EKZ (protein/RNA complex), 1A60 and 1P5M (both containing RNA).

automatic mode, in relation to the core atoms found by CYRANGE. In general, the domains identified by GeoStaS that had the smallest R_{division} were similar to the ones found by CYRANGE, but in some cases there were discrepancies. GeoStaS found too many domains for very small NMR ensembles (up to 12 conformations) of structures having flexible loops or termini. One such case is depicted and described in Figure 4.4. This unreasonable division produced by GeoStaS is most likely caused by not enough data in the ensemble for our algorithm, which is optimized for a larger number of input conformations. As can be seen from Equation 4.6, the R_{division} value is based on averaging, thus the more data is provided, the more precise the division. Overall, for proteins GeoStaS showed a similar domain division as the CYRANGE software and the smaller value of R_{division} pointed to a more reasonable division.

For the tests on experimental data of nucleic acids, we chose several ensembles that had more than 20 conformations. The results of the analysis of two ensembles are presented in Figure 4.5. The ensemble 1P5M[197] contains 36 conformations of the HCV IRES domain IIa — a non-canonical RNA helix with a large internal loop region, formed by five bases (A53, A54, C55, U56 and A57). The division found by GeoStaS shows that the loop and the flanking

helical regions were identified as separate domains (Figure 4.5A–C). This agrees with the experimental analysis, which demonstrated that the whole structure was bending in relation to the loop region.[197] Another ensemble, 1EKZ,[198] is a complex of a double stranded RNA with a double-stranded RNA-binding domain from an *E. coli* protein. The visualization of the GeoStaS analysis, shown in Figure 4.5D–E, illustrates that only a small part of the protein "moved along" with RNA (grey domain), so as to maintain the hydrogen bonds that were shown to contribute the most to a stable complex.[198]

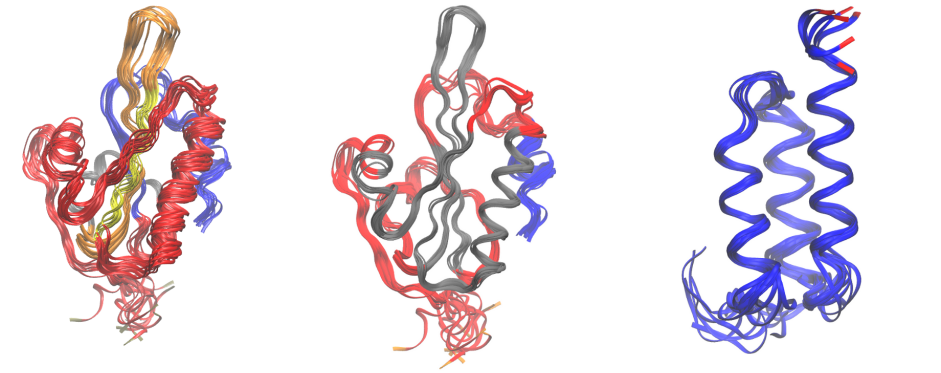


Figure 4.4: **NMR ensembles with few conformations can be problematic for GeoStaS**, if they represent a very flexible molecule. The optimal division seems too complicated *(A)* but when setting the threshold to a higher value, a more reasonable division intro dynamic domains is obtained *(B)*. Another NMR ensemble that has few conformations was not as problematic *(C)* because the conformations did not differ much from each other.

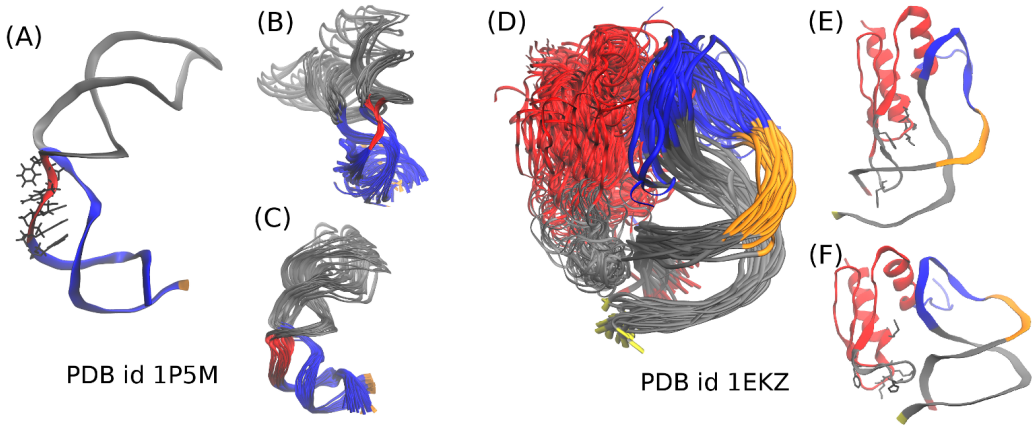


Figure 4.5: **NMR ensembles containing nucleic acids, colored by the domains recognized by GeoStaS (Cα/P representation, automatic mode).** *(A–C)* RNA non-canonical helix: *(A)* one conformation with the loop-forming bases marked as black sticks, and all NMR conformations superimposed with respect to the red *(B)* or blue *(C)* domain. *(D–E)* Protein:RNA complex: *(D)* all conformations superimposed with respect to the grey domain, *(E)* two chosen conformations illustrating how the protein "follows" the nucleic helix. Amino acids whose mutations resulted in abolishing of RNA binding[198] are shown as sticks.

Simulations

We aimed at facilitating the analysis of conformational changes in biomolecular simulations, and therefore the main tests were conducted on various trajectories obtained from MD simulations. The first test was performed on the simulation of a GroEL chaperone monomer complexed with ATP (Figure 4.6).[199] The chaperonin GroEL is a large oligomeric protein (ca. 800 kDa) that helps fold non-native proteins in *E. coli*.[200] After forming a complex with the co-chaperonin GroES, it provides a protective chamber, where the substrate proteins regain their functional shape. During this complexation the subunits of GroEL undergo large conformational changes, which have been extensively characterized.[199,201,202]

This analysis required using GeoStaS in the manual mode, probably due to the large size of the system and the complexity of the movements. However, setting the threshold to 0.05 gave an interesting division into dynamic domains, as presented in Figure 4.6B–C. Structural studies divide the GroEL monomer into three domains:[201] equatorial, intermediate, and apical, as illustrated in Figure 4.6A. The division found by GeoStaS resembles these structural domains, however, the apical and equatorial domains were subdivided into smaller dynamic domains: yellow and gold fragments in the apical part, and blue, grey and red in the equatorial part. The yellow-colored dynamic domain contains helices K and L, which stay close to each other during the elevation of the apical domain while forming the chamber in the GroEL chaperone. Therefore, their movements should be slightly different than the movement of the rest of the apical domain. Moreover, the red-colored part of the equatorial domain consists of helices that interact with the bound ATP, which again would cause their movements to differ from the rest of the equatorial domain. This is again in good agreement with previous structural studies highlighting the internal motions in the equatorial domain upon ATP binding.[199,203]

We compared the similarity coefficients from our AMSM with the standard DCCM values and we found some discrepancies. Figure 4.7 highlights some areas that differ between these matrices. For example, AMSM showed correlations of motions of helices A, B and C with helix M and their anti-correlations with helices K and L, contrary to the DCCM analysis (area 1 in Figure 4.7). The four former helices (A, B and C from the equatorial domain, and M from the intermediate domain) move in the same direction during the first stage of conformational changes in the GroEL cycle,[202] while helices K and L, situated in the apical domain, move in an opposite direction during this transition. Therefore, the relations found by AMSM seem reasonable. Moreover, helices F and G were anti-correlated with helices Q and R in DCCM, in contrast to AMSM (area 2). It has been shown that first, helices F and G (from the apical domain) approach the equatorial domain (thus also helices Q and R), in order for the apical domain to move upwards and then to a fully opened conformation.[202] Similar case is between helices N, O and helix F (area 3). Again, the relation found by AMSM seems justified.

Another test of a protein simulation was conducted on the trajectory of a bacterial elongation factor, EF-Tu[204] (Figure 4.8). This protein complexates with GTP, and subsequently with aminoacylated tRNA, to deliver it to the ribosome. Next, thanks to the energy that comes from the GTP→GDP hydrolysis, the tRNA is released and the EF-Tu:GDP complex dissociates from the ribosome. Both, the experimental[205–207] and computational studies[204] showed that some parts of the protein were significantly more dynamic when GDP was bound than without the nucleotide diphosphate — these fragments are termed switch I and II. Our analysis correctly identified switch I as a distinct domain (colored orange in Figure 4.8B). In addition, the majority of switch II residues were assigned to a fragment of the red domain, which suggests that its movements differed from the movements of the neighboring residues, which were in the gray domain.

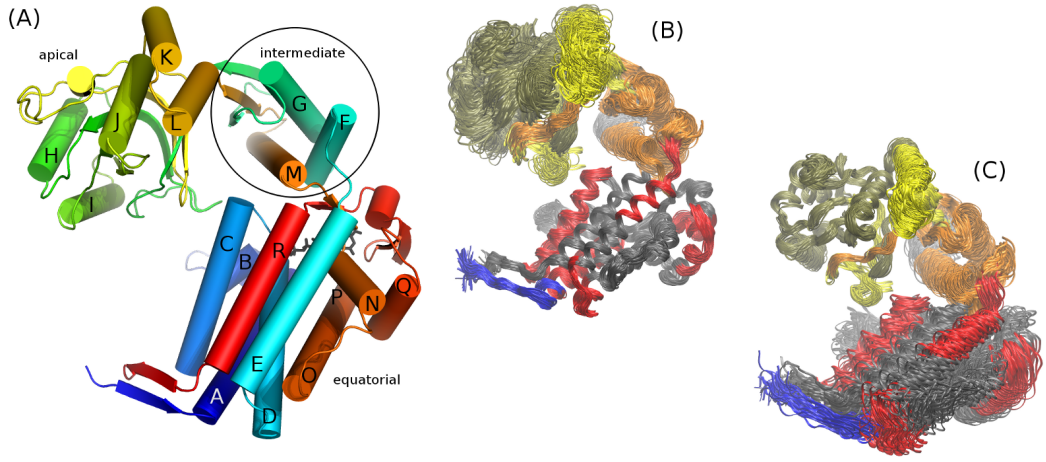


Figure 4.6: **Division into structural and dynamic domains of the GroEL monomer.** *(A)* A cartoon model of the monomer with marked helices and structural domains. *(B–C)* Visualization of the domains found by GeoStaS for the threshold (Δw) set to 0.05 — several trajectory conformations are superimposed relative to the domains colored: *(B)* grey, and *(C)* gold.

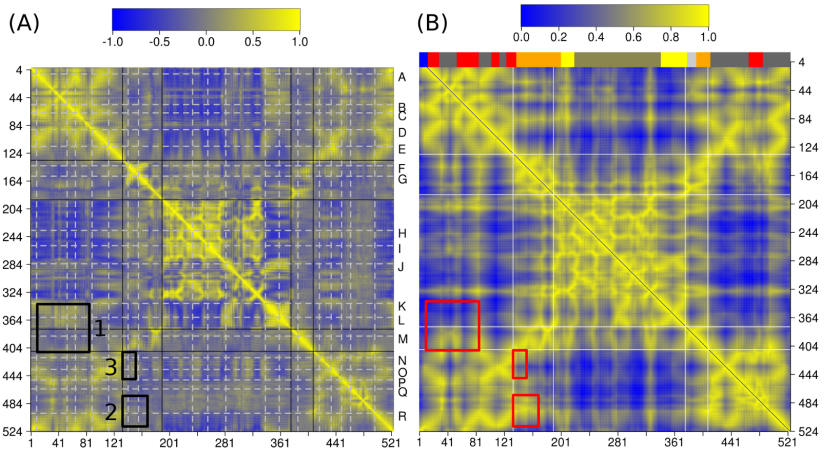


Figure 4.7: **Analysis of the correlation patterns in the GroEL monomer — *(A)* DCCM versus *(B)* AMSM**; some of the areas that differ are marked by the red and black squares and discussed in the text; the meaning of the color scales is the same in both matrices: dark blue shows strong anti-correlations, and yellow — strong correlations. Above the AMSM, the color bars represent different dynamic domains (coloring is the same as in Figure 4.6B–C). The letters to the right of DCCM mark the GroEL helices shown in Figure 4.6A.

We compared the AMSM and DCCM values for the EF-Tu:GDP simulation, and additionally, we applied the clustering algorithm to both matrices, in order to compare also the resulting divisions into dynamic domains (Figure 4.9). Again, we noticed differences between these matrices. For example, the area marked 1 in Figure 4.9 shows the relation between movements of switch I and domain II of EF-Tu (see Figure 4.8A), which was identified as correlation by DCCM and anti-correlation by AMSM. Moreover, the DCCM-derived dynamic domain colored blue contained switch I together with domains II and III. As described above, switch I was found to be much more mobile than the rest of the protein, therefore the AMSM result seems more reasonable. The relation between the movement of a fragment of domain I and domain III is enclosed in the area marked 2. The original analysis of the simulation[204] showed that domain I moved away from domain III, and the anti-correlation found by AMSM agrees with this observation. Finally, the area number 3,

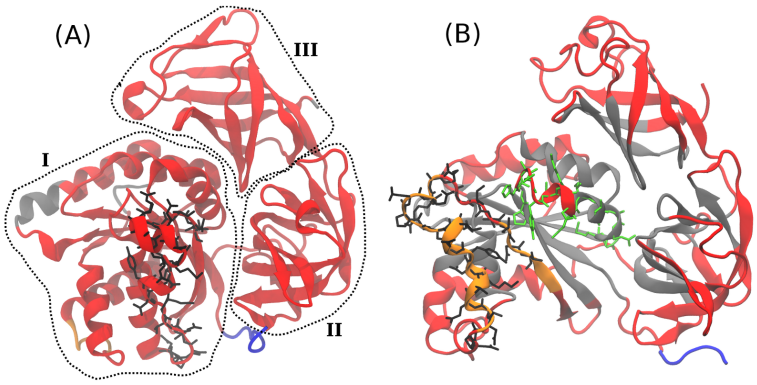


Figure 4.8: **Division into dynamic domains of** *(A)* **EF-Tu (first trajectory frame) and** *(B)* **EF-Tu:GDP complex (last trajectory frame; ligand is not shown for clarity).** Each structure is colored independently. Structural domains are encircled and numbered in panel *(A)*. Residues that form switch I are depicted as sticks and colored black in both panels; in panel *(B)* also switch II is marked, colored green.

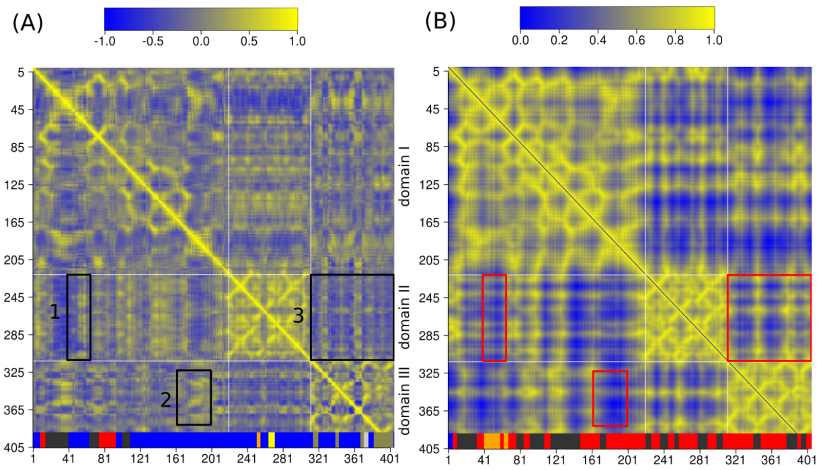


Figure 4.9: **Analysis of the correlation patterns in the EF-Tu:GDP complex —** *(A)* **DCCM versus** *(B)* **AMSM**; some of the areas that differ are marked by the red and black squares and discussed in the text; the meaning of the color scales is the same in both matrices: dark blue shows strong anti-correlations, and yellow — strong correlations. Below the matrices, the color bars represent different dynamic domains (the coloring below AMSM is the same as in Figure 4.8B).

according to AMSM, points to a correlation between the movement of domains II and III, but according to DCCM this movement was anti-correlated. Additionally, DCCM-derived division into dynamic domains classifies domain III to a different dynamic domain than the majority of domain II (yellow and blue, respectively). Again, the AMSM result is in accord with the original analysis, where the researchers reported that these fragments did not recede.[204]

Next, we performed tests on systems containing nucleic acids: trajectories of an elbow segment of helix 38 (H38) of 23S ribosomal RNA from *Haloarcula marismortui*[208] and of bacterial elongation factor EF-Tu complexed with tRNA.[204]

As mentioned above, EF-Tu, complexates with tRNA to deliver it to the ribosome during the translation process. The shape of this complex highly resembles the shape of another factor also involved in the peptide elongation process of bacterial translation, EF-G. These two structures are presented in Figure 4.10. Apart from the shape, the motions and also their binding sites on the ribosome are similar. Structurally, EF-G was divided into five domains (Figure 4.10A), but dynamically, only two domains were found by GeoStaS (Figure 4.10B–C). The red-colored domain constitutes a core formed of structural domains: I, II, III, and V. The G′ insertion, which is exclusive for EF-G, was classified as a separate dynamic domain

(colored blue) together with the flexible, solvent-exposed fragments. Almost the entire domain IV was assigned to this blue domain, which is reasonable because this domain bends upon binding of EF-G to the ribosome.

The division of the EF-Tu:tRNA complex is presented in Figures 4.10E–F. According to the GeoStaS analysis, the anticodon arm of tRNA was moving separately, while the acceptor stem together with the EF-Tu protein formed one dynamic domain (colored red). This resembles slightly the division of EF-G.

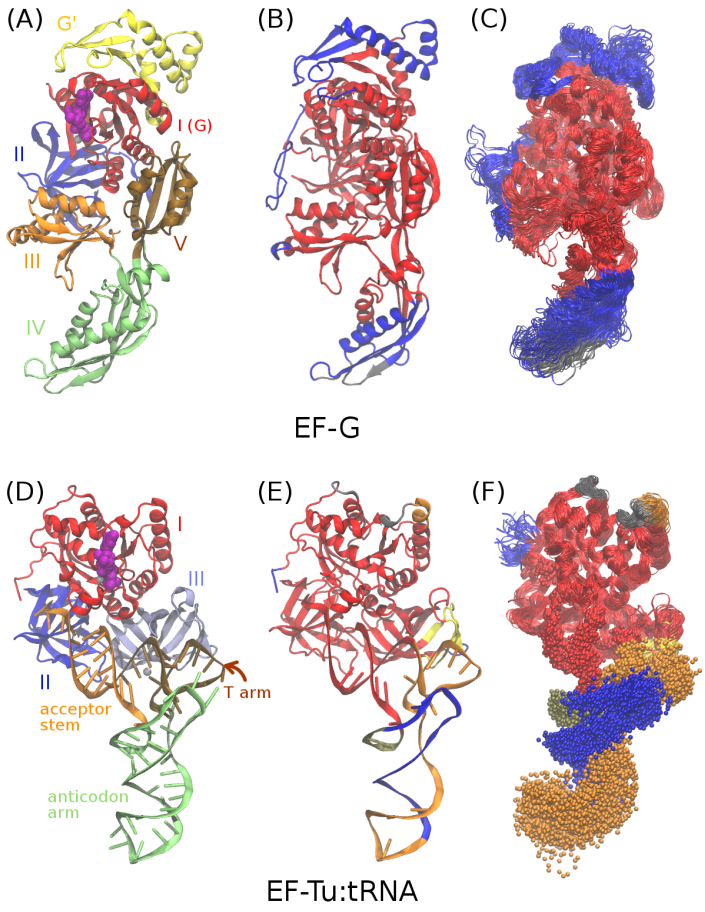


Figure 4.10: **Comparison between the structural (panels *A* and *D*) and dynamic domains (panels *B–C* and *E–F*) for the EF-G protein and the protein:RNA complex — EF-Tu:tRNA.** In panels *(C)* and *(F)* several conformations from MD simulations are superimposed relative to the red domain. The division of EF-G was calculated for the Cα representation (automatic mode), while for the EF-Tu:tRNA the presented division was for the three-atoms-per-residue representation (automatic mode).

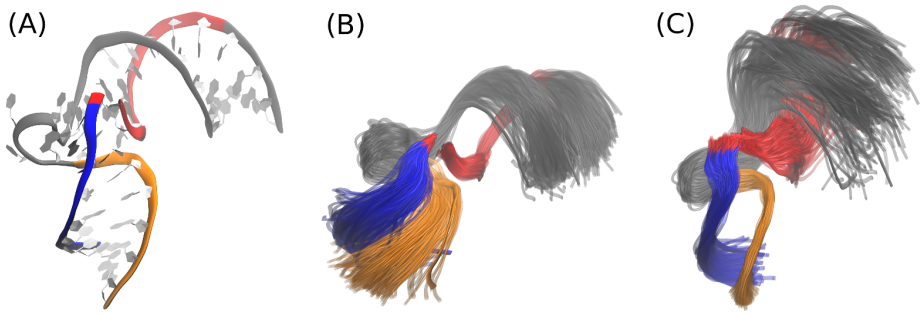


Figure 4.11: **Division into dynamic domains found by GeoStaS for the simulation of H38 fragment.** *(A)* Initial structure of the H38 kink-turn region; the backbone is colored according to the optimal division into dynamic domains found by GeoStaS with the one-atom-per-residue representation. Snapshots from the simulation are superimposed with regard to the domain colored *(B)* red or *(C)* orange, to emphasize the movements observed in the trajectory.

The last case presented here is the analysis of the simulation of a H38 RNA fragment from the bacterial ribosome.[208] This RNA fragment forms a kink-turn motif, which is commonly found in ribosomal RNA structures. The kink-turns are very dynamic, often initiating large conformational changes.[209] The dynamic domains found by GeoStaS show exactly this high flexibility, as visualized in Figure 4.11. The H38 kink-turn fragment is divided into four parts, which emphasizes two types of motion: the bending mode, which changes the angle between the flanking helices, and the breathing mode of the helical parts. The same conclusions were drawn in the original paper,[208] albeit only after many measurements of distances and angles between the two strands. Our analysis simplifies this process.

4.1.3 Summary

We have presented an algorithm, GeoStaS, for dividing a molecule into dynamic domains based on its conformations. The novelty of this algorithm is its focus on geometric comparison of the atom traces, which enables analysis of any set of conformations, regardless of their order. Additionally, this approach properly recognizes all types of movement similarities, including rotational correlations. The algorithm was implemented in a graphical software, which has a minimal number of parameters and automatically suggests an optimal solution. We have shown that with GeoStaS the analysis of internal molecular motions can be simple. The software deals well with large datasets from simulations of various biomolecules, including proteins, nucleic acids and their complexes. The resulting division into dynamic domains helps describe the changes between subsequent conformations. We showed that GeoStaS is able to identify the correlated vs. anti-correlated motions better than the standard cross-correlation matrix, because of accounting also for rotational correlations.

The algorithm presented here is generic because it can deal with any objects that are comparable through geometric transformations. Therefore, not only conformations of one molecule can be analyzed but also, for example, conformations of its mutated variants. Moreover, the outcome of the analysis can have different applications, such as for generating different conformations for flexible docking procedures or fitting a model into low resolution microscopy maps.

4.2 Bacterial resistance mechanism 1: ribosomal RNA A-site mutations

The two following sections describe the part of the research that focused on the two bacterial resistance mechanisms, namely the modifications of bases in the primary aminoglycoside binding site, the 16S rRNA A-site, and the production of the enzymes that chemically alter aminoglycosides. First, we describe the results from our simulations of the A-site models with different mutations, with and without paromomycin.

4.2.1 Conformations of bound paromomycin

To investigate the conformations that paromomycin adopts when bound to the mutated A-site, we calculated the ϕ/ψ dihedral angles of paromomycin's linkages, as described in Section 3.3.2 *Parameterization validation*. Figure 4.12 presents the ϕ/ψ conformations of paromomycin in different A-site models. In the WT A-site, paromomycin adopted a conformation that was also the most frequently sampled one in solution (compare Figure 4.12, top, with Figure 3.4E–G). Moreover, this conformation did not change when paromomycin complexated with the mutated RNA structures. Upon binding to different rRNA molecules, the mobility of all three linkages of paromomycin was only restrained to the region most populated in water. The most flexible part of paromomycin was the ring IV, which protrudes from the A-site. Only in the G1491A and U1406C/U1495A simulations, this ring was slightly more mobile than in the other simulations of the complexes.

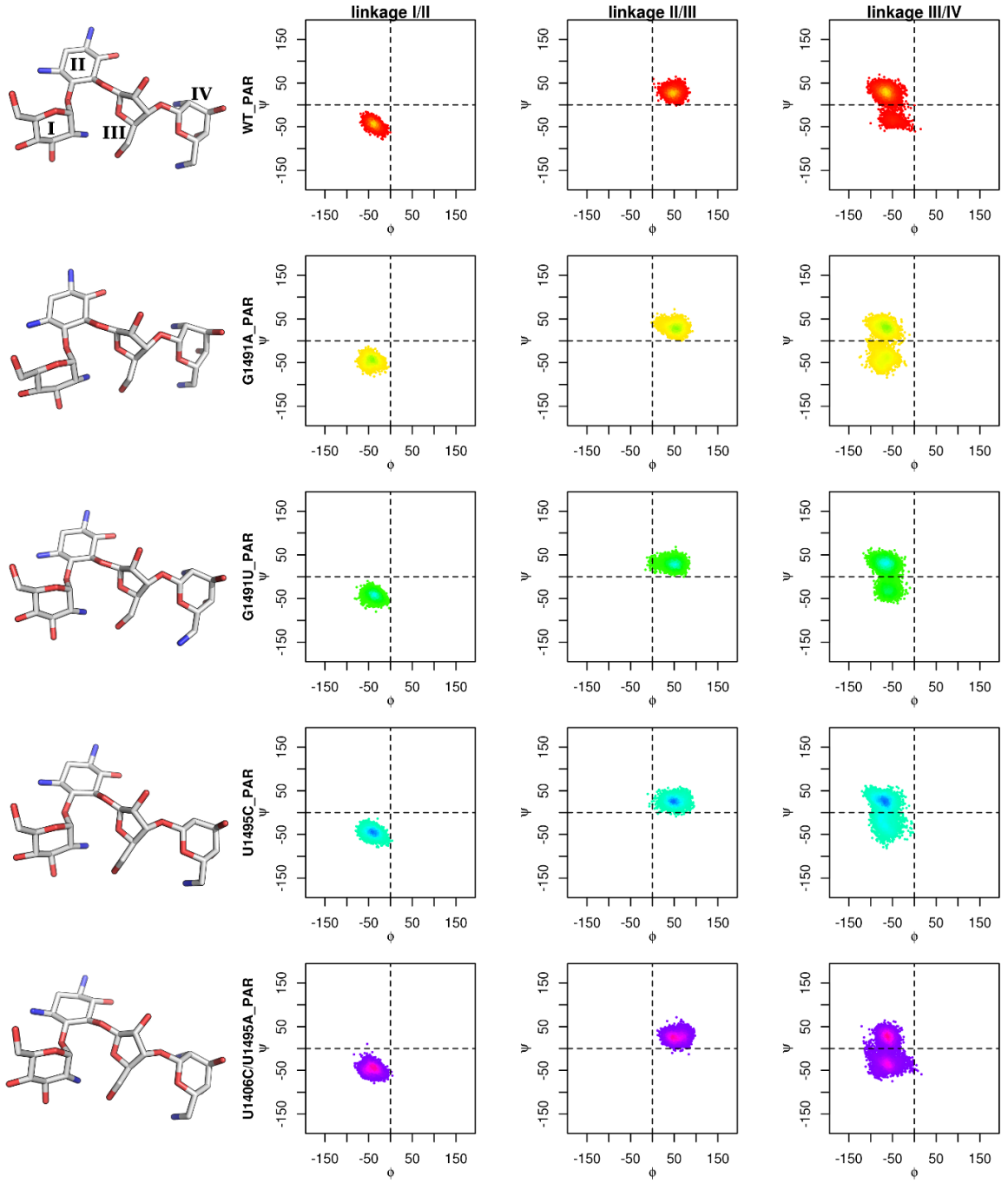


Figure 4.12: **Conformations of paromomycin bound to different RNA receptors.** On the left, the most populated conformations from the simulations are presented. The graphs depict changes of the dihedral angles of the linkages between the rings.

4.2.2 Internal dynamics of the bare ribosomal RNA A-site

To check the mobility and stability of the simulated RNA models, first, we looked at some basic measures, i.e., RMSD of the atomic positions from the initial conformation and RMSF of each residue. In all simulations, RMSD of the RNA A-site was leveled and the average did not exceed 2.9 Å (Table 4.1). These measures revealed an overall stabilization of the A-site upon aminoglycoside binding, although at a substantially smaller level in the simulation of the doubly-mutated A-site, i.e., U1406C/U1495A vs. U1406C/U1495A_PAR. This was also visible in the range of RMSF values, as shown in Figure 4.13. In the WT_PAR simulation, there was a shift towards smaller values compared to WT. Such a trend was also observed in the simulations where G1491 was mutated. On the contrary, for the simulations where the U1406·U1495 pair was mutated, the histograms overlap. These findings are in agreement with the results from previous simulations of different aminoglycosides bound to RNA, which showed that the WT complexes were more stable than the free A-site.[85,86]

Table 4.1: **Average RMSD and ligand RMSF for the RNA simulations**

	avg RMSD [Å][(a)]		**ligand RMSF [Å]**	
simulation name	**part A**	**part B**	**part A**	**part B**
WT[(b)]	1.56±0.26	2.35±0.39	—	—
WT_PAR	1.30±0.22	1.41±0.17	0.86	0.74
WT_KAN	1.69±0.28	1.86±0.23	1.10	1.10
G1491A	2.65±0.34	2.72±0.45	—	—
G1491A_PAR	1.71±0.36	1.71±0.18	2.37	0.70
G1491U	2.69±0.42	2.83±0.46	—	—
G1491U_PAR	1.37±0.21	1.51±0.20	0.97	0.86
U1495C	2.38±0.24	2.71±0.34	—	—
U1495C_PAR	1.79±0.23	1.91±0.30	2.09	1.14
U1406C/U1495A	2.42±0.27	2.15±0.40	—	—
U1406C/U1495A_PAR	2.18±0.60	1.60±0.23	3.00	1.59

[(a)] Average values shown with the standard deviation, calculated for all heavy atoms in each of the symmetrical parts of the model, labeled A and B (see Figure 3.1);
[(b)] Data taken from our previous study.[85]

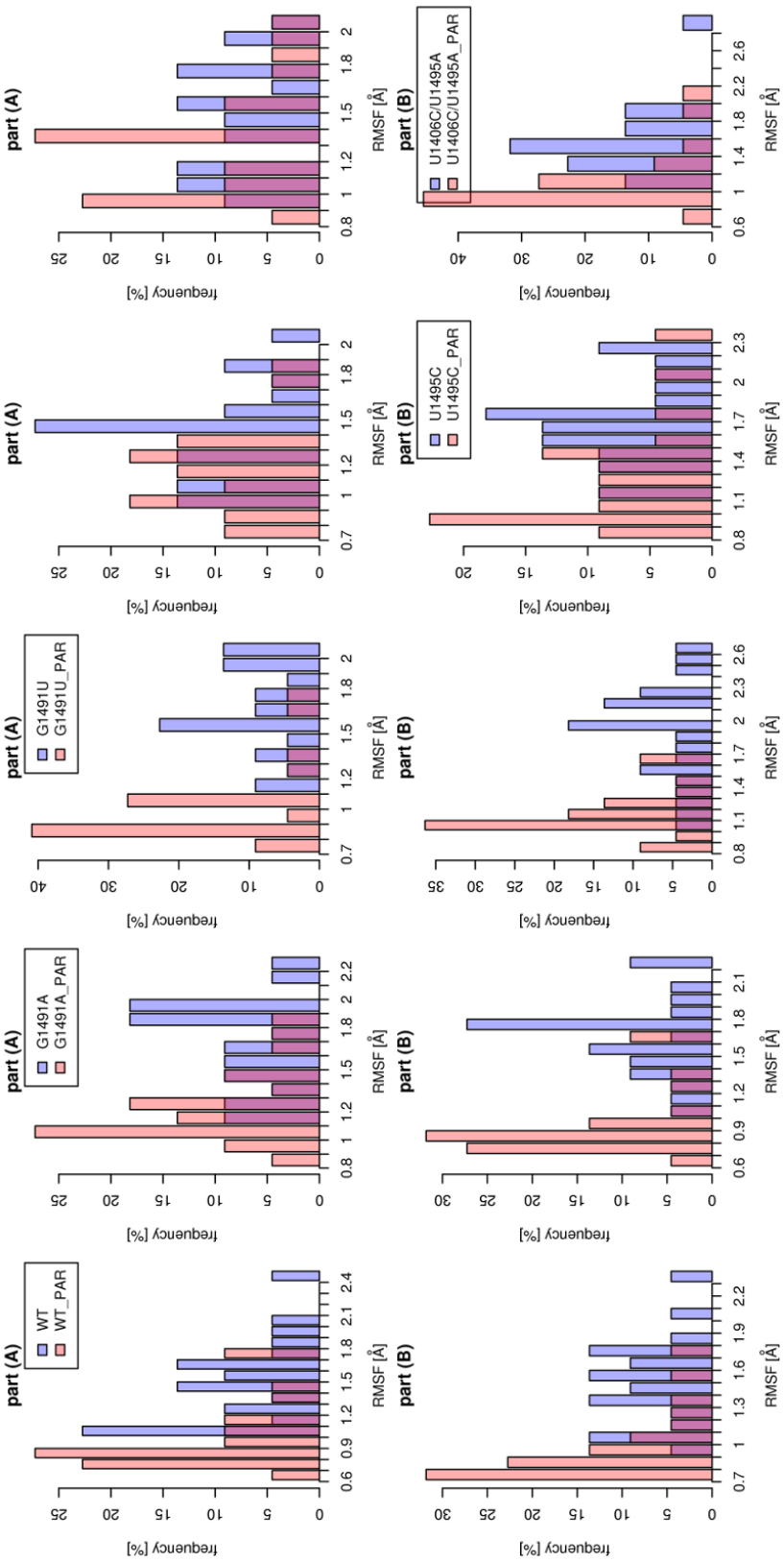


Figure 4.13: **Histograms of per-residue RMSF values [Å] in the MD simulations of the A-site models.** Each column shows values for different simulations (WT and the studied mutants, with and without paromomycin), and the two rows depict values for the two symmetrical parts of the A-site model.

Monitoring hydrogen bonds in base pairs

To investigate the source of the observed differences in the RMSF ranges (Figure 4.13), we monitored each of the bases and base pairs in the simulations of the mutated structures without paromomycin. Figure 4.14 depicts the number of hydrogen bonds formed within each base pair for three simulations: WT, G1491A, and U1406C/U1495A. We found that some base pairs were unstable in the simulations G1491A, G1491U, and U1406C/U1495A. Specifically, it appears that the double mutation did not provide enough stability for the newly formed 1406C:1495A base pair, while the mutations of G1491 were causing disruption of its pairing with the opposite RNA strand. In addition, we noticed that the bulged adenines, A1492 and A1493, formed hydrogen bonds with the opposite A1408 more often in the structures with mutated base G1491 than in the remaining ones. These observations are discussed in details below.

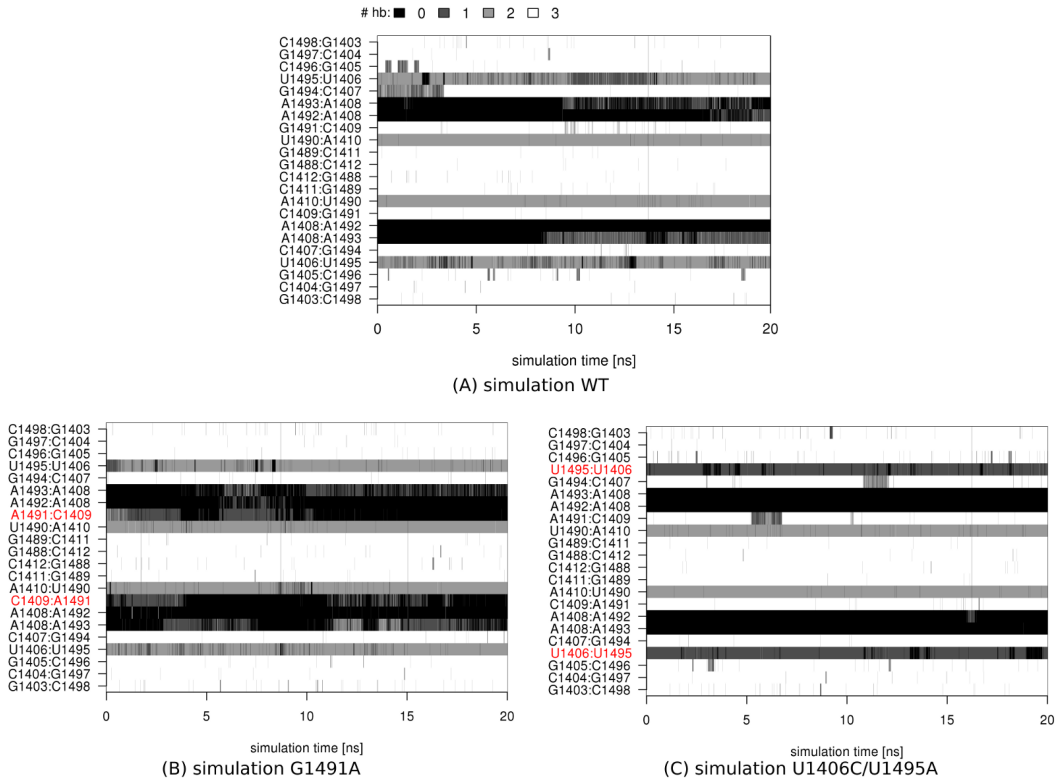


Figure 4.14: **The number of hydrogen bonds in base pairs versus simulation time** for selected simulations. The base pairs with mutated bases are marked red and the number of hydrogen bonds is depicted in grayscale code.

Stability of the U1406·U1495 pair is disrupted upon mutations

Previous studies have suggested that the stability of the U1406·U1495 base pair influences the stability of the entire A-site.[45,84] Our simulations suggest that the double mutation (U1406C/U1495A) significantly interfered with the stability of this base pair. In the X-ray structure of the WT A-site (1J7T), the U·U pair is formed by two hydrogen bonds, which was also reproduced in our WT simulation, as shown in Figure 4.14A. In the doubly-mutated A-site, the resulting C:A pair shared maximally one hydrogen bond (Figure 4.14C). The structures with neither of the uridines mutated preserved the wild type-like geometry of the U1406·U1495 pair with two hydrogen bonds, as presented in Figure 4.14B for the G1491A

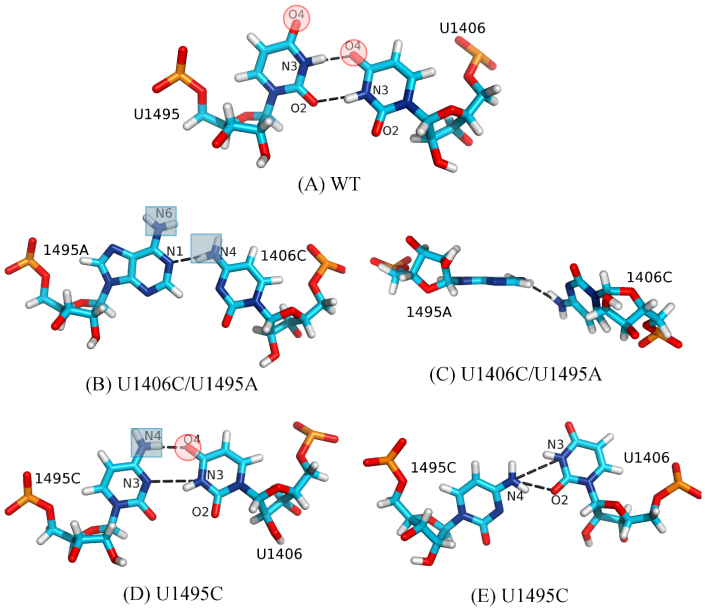


Figure 4.15: **Conformations of the U1406·U1495 pair and its mutated alternatives** observed in the simulations of the bare A-site models. The distribution of charges changed upon mutations — negatively charged moieties are marked with red circles, positively charged ones are marked with blue squares. [Figure taken from our published results[211]]

simulation. In the U1495C structure, the resulting base pair (U1406:1495C) was formed with two or even three hydrogen bonds (data not shown). In summary, bases 1406 and 1495 formed a pair for the majority of the simulation time in all simulations (Table 4.2); however, the geometry and hydrogen bonding strength of this pair varied. Figure 4.15 compares the different base pairing observed for U1406·U1495 in the WT and mutated structures. In the WT simulation, the U1406·U1495 base pair adopted a stable conformation with two hydrogen bonds (Figure 4.15A). The double mutation (U1406C/U1495A) resulted in a less stable C:A pair, which often adopted an experimentally observed conformation[†] (Figure 4.15B). However, sometimes the mutated 1406C base rotated to an almost perpendicular position relative to the other base (Figure 4.15C), which indicates a weak hydrogen bonding. We did not observe such a behavior in any other simulation. For the majority of the U1495C simulation time, the conformation of the U:C pair resembled a well-known 4-carbonyl-amino pattern[127] (also called *cis* W.C./W.C.;[210] depicted in Figure 4.15D). Sometimes a different hydrogen bonding pattern occurred, as illustrated in Figure 4.15E.

In the U1406C/U1495A structure, upon the change of base 1495 from a smaller pyrimidine (uracil) to a larger purine (adenine), the shape of the base pair was disrupted and during the simulation the adenine was situated more towards the aminoglycoside binding site than the uracil in the wild type structure. This change in the 1406C·1495A pair geometry may hinder paromomycin binding through steric clashes. However, it did not influence the overall stability of the A-site model, which is in agreement with previous study.[34] Only the mobility of the A1492 and A1493 has been affected, as discussed below.

Importantly, also the charge distribution was altered upon mutations within the U1406·U1495 pair, as can be seen in Figure 4.15. In the WT structure, these two uracils provide negatively charged C=O moieties that form hydrogen bonds with aminoglycosides. The double mutation deletes this negative potential completely, while the U1495C substitution provides only one of the original moieties. Therefore, even if the geometry of the base pair was not altered, the change in the charge distribution could hinder forming of hydrogen bonds with an aminoglycoside binding to this site.

[†]As reported in the Database of RNA base pair structures: http://bps.rutgers.edu/atlas/bppattern/ac_5

Substitutions of base G1491 cause a shift in base-pairing

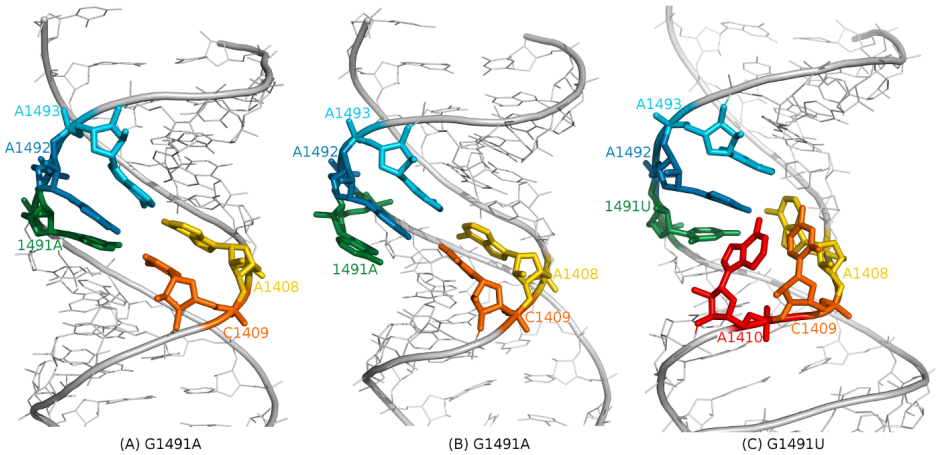
Another disruption in base pairing pattern involves G1491 in the simulations where it was mutated to either 1491A or 1491U. Table 4.2 compares the duration of base pair forming for the relevant simulations. During the WT simulation, the G1491:C1409 base pair was stable, forming three hydrogen bonds in both parts of the structure almost throughout the simulation time (see Figure 4.14A). Introducing the mutation resulted in a poorly bound pair throughout the G1491A simulation, as was shown in Figure 4.14B, and even less stable in G1491U. The visualization of trajectories showed a shift in the base pairing pattern (Figure 4.16), where some bases were in the flipped out position (1491A in Figure 4.16B or C1409 and A1410 in Figure 4.16C). Due to the intrinsic flexibility of RNA such changes occur relatively often[212] and we observed a similar shift in our previous study.[85] Results presented herein suggest that the different base pair pattern resulted in a change of the binding cleft's shape (described below).

Table 4.2: **Base pairing in the RNA simulations**

		duration of pairing [%]									
base pairs ↓		**G1491A**		**G1491U**		**U1495C**		**U1406C/U1495A**		**WT**[a]	
1406 :[b]	1495[c]	100	99	98	99	81	84	89	90	96	97
A1408 :	A1492	6	1	4	2	0	15	2	0	0	9
	A1493	49	36	47	25	6	24	0	0	50	25
C1409 :	1491[d]	29	34	16	21	*not applicable*				100	100
	A1492	32	16	18	10	*not applicable*					

Percentage of simulation time when the base pairs were formed (i.e., at least one hydrogen bond was present). Two values are shown for each simulation corresponding to two A-sites in the model.

[a] data from our previous study;[85]
[b] this base is U in G1491A, G1491U, WT and U1495C; and C in U1406C/U1495A simulation;
[c] this base is U in G1491A, G1491U and WT; C in U1495C; and A in U1406C/U1495A simulation;
[d] this base is A in G1491A, U in G1491U, and C in the other simulations.



(A) G1491A (B) G1491A (C) G1491U

Figure 4.16: **Shift of base pairing pattern observed in some RNA simulations**.

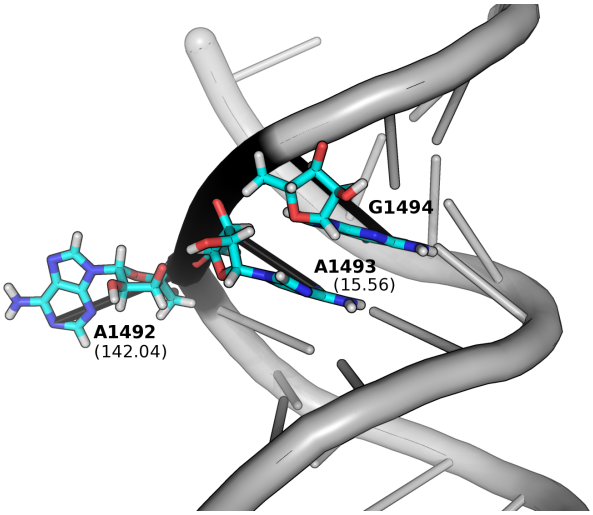


Figure 4.17: **Pseudo-dihedral angles describing the variation in the conformations of nucleic bases A1492 and A1493**. The angle θ_{A1493} is calculated as a torsion angle between the four atoms: G1494(N1) – G1494(P) – A1493(P) – A1493(N1), as depicted by black sticks (analogously for θ_{A1492}: G1494(N1) – G1494(P) – A1492(P) – A1492(N1)). The exemplary values of the pseudo-dihedral angles (in degrees) are shown in brackets. [Figure taken from our published results[211]]

Mutations influence the mobility of A1492 and A1493

In the wild type A-site, the most mobile bases are A1492 and A1493.[11,84,85] These adenines are in a dynamic equilibrium between the flipped out and flipped in conformations, and they form hydrogen bonds with the opposite A1408. We described the relative conformations of A1492 and A1493 by defining pseudo-dihedral angles, relative to the succeeding base in the sequence, G1494 (Figure 4.17). This method of quantifying the mobility of nucleic bases has been used before.[83,213–215] Thus measured different relative orientations of the two adenines can be organized in three basic states:[20–27]

(A) both adenines occupy the inside of the RNA helix ($\theta_{A1492} \in [-45; 50]$, $\theta_{A1493} \in [-45; 40]$); this conformation prevents binding of an aminoglycoside and indicates a rejection of a non- or near-cognate tRNA during a normal translation process (Figure 4.18A);

(B) A1492 is flipped out ($\theta_{A1492} < -45$ or $\theta_{A1492} > 50$) and A1493 occupies the inside of the helix ($\theta_{A1493} \in [-45; 40]$); the recognition of the translation termination factor requires this conformation (Figure 4.18B);

(C) both A1492 and A1493 are flipped out ($\theta_{A1492} < -45$ or $\theta_{A1492} > 50$, and $\theta_{A1493} < -45$ or $\theta_{A1493} > 40$), which indicates the acceptance of a cognate tRNA and also the easiest access for an aminoglycoside (Figure 4.18C).

We checked whether the mutations affected the mobility of these bases. Figure 4.18 shows the sampling of the conformation space of A1492 *versus* A1493, observed in different simulations. In the WT simulation, the adenines were dynamic and visited all the mentioned conformations. Compared to the WT simulation, the largest changes occurred in the simulation with the doubly mutated U1406·U1495 pair — in this case, the adenines were almost never inside the RNA helix. On the contrary, the motion of A1492 and A1493 was restricted to the flipped in conformations in the simulations G1491A and G1491U. Interestingly, the least changes were introduced by the U1495C mutation. This can be also observed through measuring the percentage of time when A1492 or A1493 occupied the flipped-in state (Table 4.3). These adenines were inside the helix for a significantly shorter time in the U1406C/U1495A simulation than in the WT. A similar trend was also observed for one part of the model in the U1495C simulation, while for the two simulations with mutated G1491, the occupancy of the flipped-in state was almost twice as large as in the WT simulation.

Both the G1491A and G1491U mutations were found to confer high levels of resistance to paromomycin experimentally[45,47] (see Table 3.1). Our simulations showed that the occu-

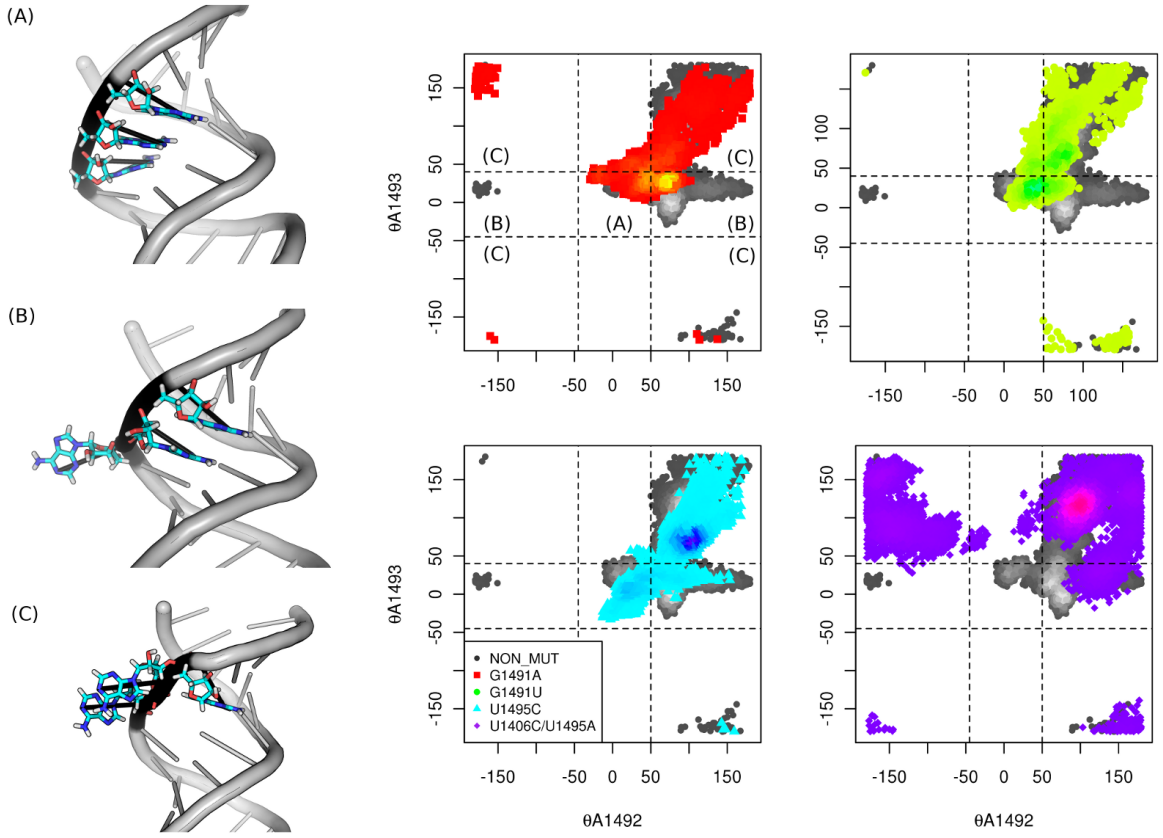


Figure 4.18: **Mobility of A1492 and A1493 in the MD simulations** quantified by the pseudo-dihedral angles, θ_{A1492} and θ_{A1493} (see Figure 4.17 for the definition). The color shading in the plots reflects the density of the points. Panels *(A–C)* show exemplary trajectory snapshots for each of the states described in the text. [Figure taken from our published results[211]]

pancy of the flipped-in state of A1492 and A1493 increased upon these mutations, compared to the WT simulation. The increased occupancy could lead to steric clashes with paromomycin, which in general can have even larger effect than the alteration of electrostatic potential in the binding site.[14] We also found that in the G1491A simulation, the adenines were moving together most of the time (compare the area of *(A)* and *(C)* conformations with the area of *(B)* conformations for the G1491A and WT simulations in Figure 4.18). The conformation *(B)*, i.e., with A1492 in the flipped-out and A1493 in the flipped-in state, is required when a translation termination factor has to be recognized and accepted.[22,23] Therefore, the change in the A1492 and A1493 movement introduced by the G1491A mutation could increase the probability of a stop codon read-through. Although this correlation between the stop-codon

Table 4.3: **Duration of the flipped-in conformations of A1492 and A1493.**

	G1491A		G1491U		U1495C		U1406C/U1495A		WT	
A1492	65	46	74	55	0	42	3	1	26	20
A1493	87	79	78	51	19	58	0	21	61	79

Percentage of the simulation time when A1492 or A1493 was inside the helix. Two values are shown for each simulation corresponding to two A-sites in the model.

read-through and the G1491A mutation was previously reported,[216] our simulations provide the first mechanistic insight and propose an explanation of this event.

Moreover, A1492 and A1493 occupied the inside of the RNA helix for a longer time in the G1491A simulation than in G1491U. However, their mobility is required during the normal translation process. Thus, this restraining may lead to worse "fitness" of bacteria possessing a G1491A substitution, and explain why the mutation G1491U is more evolutionary profitable, as was proposed earlier.[51]

4.2.3 Shapes and sizes of aminoglycoside binding pockets in RNA

Mutations of G1491 influenced the shape of RNA binding site

The observed shifts of the base pairing pattern in the RNA A-site can potentially influence the volume and shape of the aminoglycoside binding site. The pocket boundaries can be easily identified in proteins, where they often form deep cavities on a surface of a globular-shaped molecule. Contrary, the binding cleft in the RNA does not have clear boundaries. Nucleic acid helices have naturally long and wide cavities, called minor and major grooves, and the area of the binding cleft depends on the aminoglycoside. Thus, to describe the A-site binding pocket, we measured several distances between the atoms pointing to the inside of the helix. Figure 4.19 shows the representative conformations of the A-site pockets, together with the relevant intramolecular distances.

In comparison with the WT simulation, the mutations of G1491 changed the shape of the cleft — it was more compact even though the bases A1492 and A1493 moved to the flipped-in state in all three simulations (WT, G1491A, and G1491U). As can be seen, this change in shape was caused by a shift of the base pairing and twisting of base 1491. In the structures with mutations in the U1406·U1495 pair, the conformations of the two adenines were mostly flipped-out and led to a more open binding cleft.

Previous experimental studies hypothesized that the bonds formed between an aminoglycoside and a G1491A mutated A-site were weaker than in the native complex due to the less nucleophilic nature of an adenine than that of a guanine.[42] However, our results point to a different explanation, where an antibiotic could have difficulty fitting into the mutated binding site because of the apparent change in the shape of the cleft.

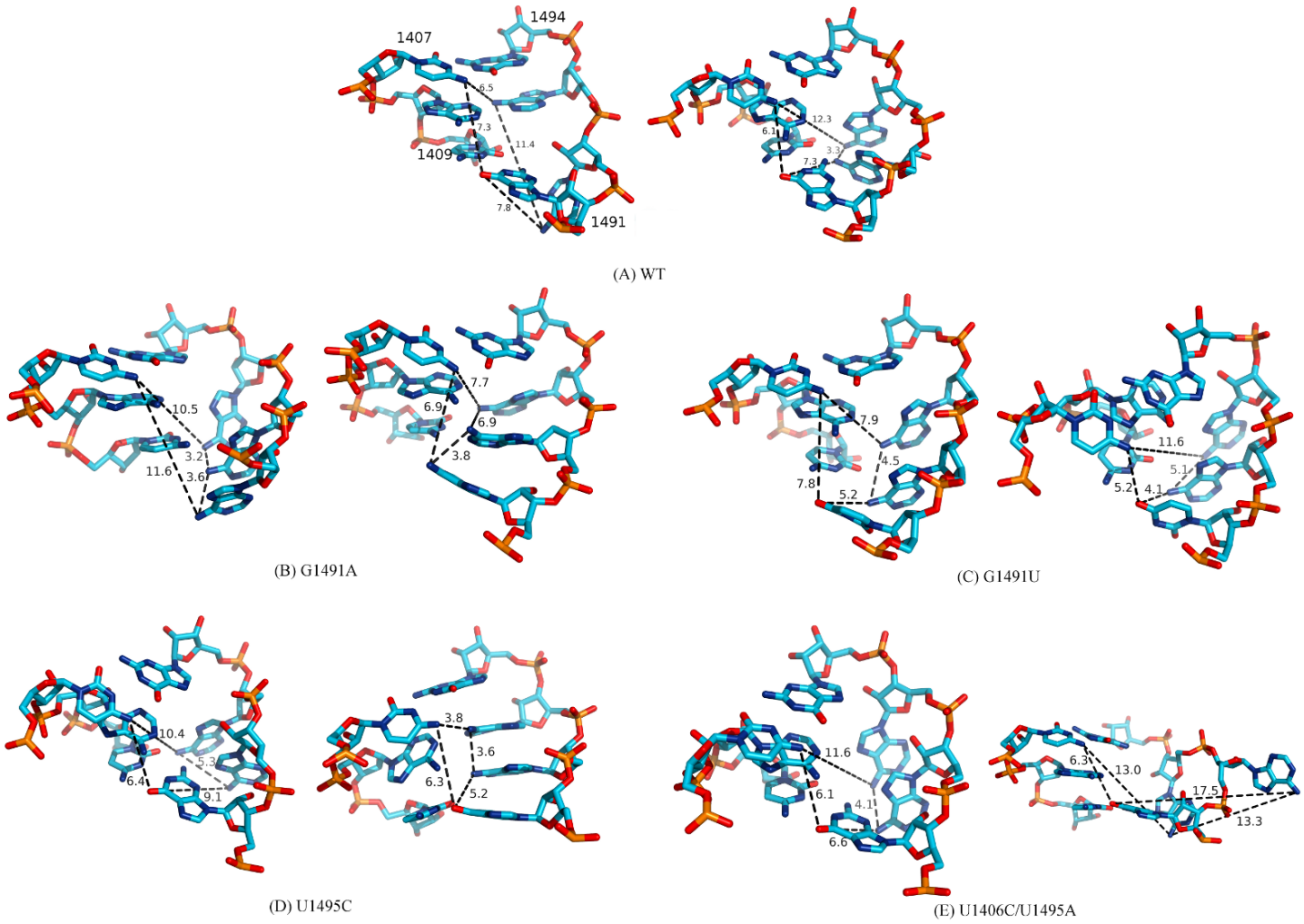


Figure 4.19: **Conformations of bases forming the A-site taken from the most populated clusters in each simulation**; the distances between the chosen atoms are labeled.

4.2.4 The U1406C/U1495A mutation changes the charge distribution inside the A-site

Aminoglycosides require a specific electrostatic potential inside a binding site because the majority of the interactions they form are of electrostatic nature.[11,217] We monitored the distribution of sodium ions in our simulations of the bare A-site models to investigate the change in the electrostatic potential of the paromomycin binding site, and to seek for any changes in the distribution upon mutations.

Previously it was observed that in the WT A-site model, the sodium ions gathered close to the phosphate groups of A1492 and A1493, and also in the position of the N3 group of paromomycin from the complex.[85] The mutations of G1491 made the ions gather slightly closer to the phosphate group of A1493 (by approximately 2 Å; Figure 4.20A–B). This indicated only a small change in the electrostatic potential, most likely caused by the two adenines being in the flipped-in state for a longer period of time than in WT type A-site. In the U1495C simulation, the area of maximal ion density was shifted $\sim$ 3 Å towards the major groove, while the double mutation U1406C/U1495A caused the sodium ions to gather outside of the binding site (Figure 4.20C–D). Since the ion density areas were not present close to the phosphate groups of A1492 and A1493, an aminoglycoside would have difficulties recognizing the doubly mutated A-site as its binding site. This corroborates with previous studies,[45,85] which showed that the negatively charged groups of the U1406·U1495 pair were important for forming a complex with an aminoglycoside.

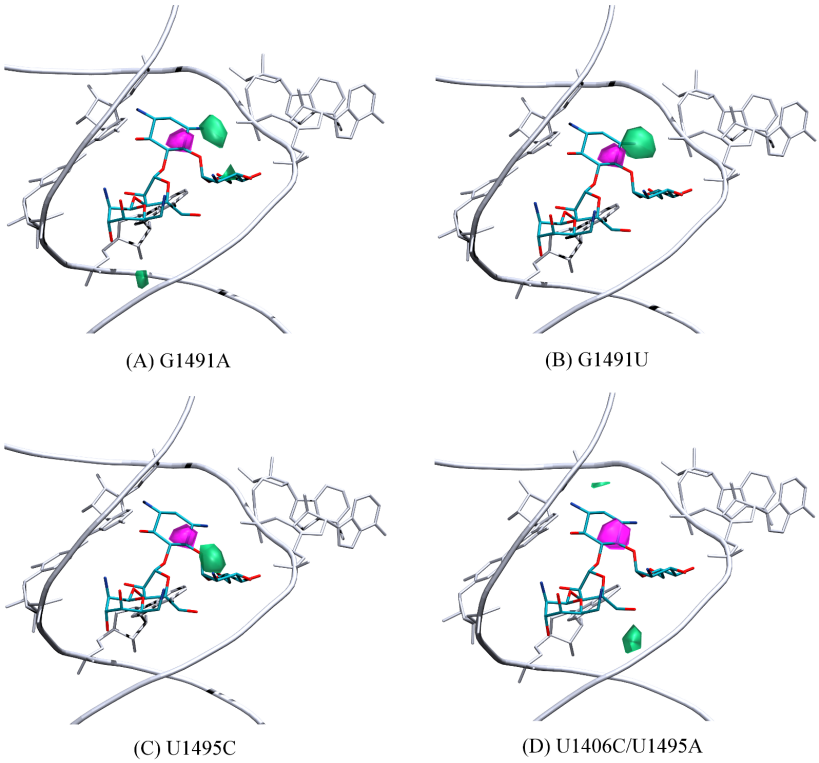
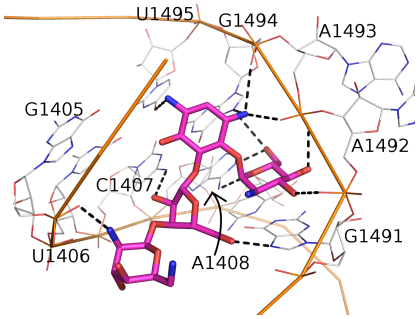


Figure 4.20: **Areas of high sodium ion density in the free A-site models:** *(A)–(C)* $\geqslant$ 0.066 ions per Å^3, *(D)* $\geqslant$ 0.055 ions per Å^3; violet areas present the data from the WT simulation, and green areas — from the simulations of the mutated structures. The superimposed position of paromomycin is shown for clarity. Only the 1406, U1495, A1408, A1492, and A1493 bases are shown in atomic details. Hydrogen atoms were not shown for clarity of the image. [Figure taken from our published results[211]]

4.2.5 Short-range interactions formed by RNA with paromomycin

To investigate the atomic details of the A-site:PAR complex formation, we analyzed and compared the interactions formed by the mutated A-site models with paromomycin. In the crystal structure of the WT A-site complexed with paromomycin, ring I of the aminoglycoside is positioned in the space created by the flipped-out A1492 and A1493, and forms a pseudo-base pair with A1408.[142] The majority of the hydrogen bonds in the crystal structure are formed with rings I and II. In our WT_PAR simulation all of these hydrogen bonds were maintained (Table 4.4).



atoms involved in H-bond	occupancy of H-bond [% of simulation time]		X-ray distance [Å]	
PAR atom . . . RNA atom	part A	part B	part A	part B
O3′. . .A1492(O2P)	95	96	2.90	3.13
O4′. . .A1493(O2P)	96	94	2.54	2.84
O5′. . .A1408(N6)	48	41	3.16	3.22
O6′. . .A1408(N1)	94	95	2.53	2.57
N1. . .U1495(O4)	94	89	2.82	2.72
N3. . .A1493(O1P)	96	96	3.13	3.17
N3. . .G1494(N7)	84	78	2.78	2.51
N3. . .G1494(O2P)	90	34	3.12	3.17
O2″. . .C1407(N4)	58	40	2.73	3.07
O5″. . .G1491(N7)	44	65	2.71	2.74
N2‴. . .G1405(O2P)	92	92	2.96	3.18
O3‴. . .U1406(O2P)	95	90	4.42	4.35

Table 4.4: **Direct hydrogen bonds formed between paromomycin and the ribosomal RNA A-site**. Only the interactions between paromomycin and a receptor's residue that lasted at least 50% of the simulation time in at least one of symmetric parts of the receptor are shown. The location of the bases relative to paromomycin is illustrated in the left panel.

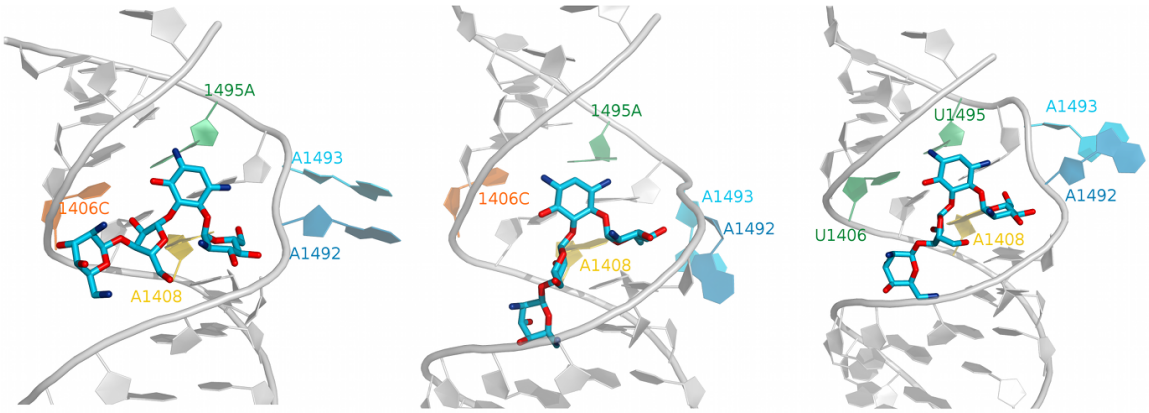
Mobility of paromomycin increases in the RNA with mutated U1406·U1495 pair

Table 4.1 shows the RMSF values of paromomycin from each of the simulations of the complexes. These values suggest that the drug's mobility was increased in the complexes with U1495C, U1406C/U1495A and G1491A mutation. The visualizations of the trajectories showed that paromomycin's position relative to the binding site changed substantially in the U1406C/U1495A_PAR simulation (Figure 4.21). We quantified these observations by measuring the distances between the centers of mass of paromomycin's rings and the center of mass of phosphorus atoms of A1492 and A1493. Figure 4.22 shows the histograms of the measured values for rings I, III and the most distant ring IV. Paromomycin was farther away from the adenines in the U1406C/U1495A_PAR simulation, in part A of the structure, as indicated by the shift of $\sim$ 2 Å in the histograms for all rings. Moderate receding of paromomycin was also visible in part B of the U1495C_PAR structure. In contrast, paromomycin in the G1491A_PAR simulation remained close to its initial position, even though its RMSF value in part A was similarly high. The measurements revealed that the drug in part B moved even closer to the adenines, as compared to WT_PAR. Similar, but less pronounced effect, was observed in part A of the G1491U_PAR simulation.

As a result of the increased mobility of paromomycin complexed with the U1406C/U1495A mutated A-site, the hydrogen bonds formed with ring II of the drug were broken. In the crystal structure 1J7T,[142] there are two important interactions between the drug and the U·U pair: *(i)* PAR(N1) forms a direct hydrogen bond with U1495(O4) (Table 4.4); and *(ii)* a water-mediated hydrogen bond is formed between PAR(O6) and U1406(O4) (the distances in the two parts of X-ray structure are 2.62 Å between PAR(O6) and the OW oxygen of

water molecules W8 or W54; and 2.59 and 2.41 Å between W8(OW) and W54(OW), respectively, and U1406(O4)). These two interactions were maintained throughout our WT_PAR simulation. Moreover, in one part of the simulated structure, a direct hydrogen bond was formed between PAR(O6) and U1406(O4) towards the end of the simulation (i.e., after ca. 11 and 14 ns in two parts of the structure, respectively). Similar behavior was observed in the G1491A_PAR simulation (Figure 4.23A–B). In contrast, in the simulation of the doubly mutated complex, paromomycin did not form any hydrogen bonds with 1495A and the interaction with 1406C was formed sporadically (Figure 4.23C–E). As shown earlier, during the U1406C/U1495A_PAR simulation, the drug changed its position relative to the binding site, thus a different hydrogen bond was formed (Figure 4.23F), which led to destabilization of the mutated base pair.

Our results demonstrated that the U1495C mutation exhibited smaller effect on paromomycin binding than the U1406C/U1495A substitution, which is in agreement with the experimental studies that showed the MIC values for bacteria possessing different mutations[44,45] (see Table 1.1). Hobbie *et al.* examined also the U1406C/U1495G mutation, which disrupts the base pair geometry and maintains only the negative charge of base 1495. Surprisingly, this substitution did not lead to significant rise in the MIC values, in contrast to the other two mentioned mutations (see Table 1.1). Therefore, it seems like the negative charge provided by the base in position 1495 is more important than the charge of the base in position 1406 or even than the geometry of this base pair.



(A) U1406C/U1495A_PAR, original conformation

(B) U1406C/U1495A_PAR, changed conformation

(C) G1491A_PAR

Figure 4.21: **Change in paromomycin's conformation in the A-site upon some mutations.** The structure with the double mutation U1406C/U1495A did not form a stably bound complex (compare *(A)* with *(B)*), but also in one part of G1491A mutated A-site the drug changed its conformation *(C)*. [Figure taken from our published results[211]]

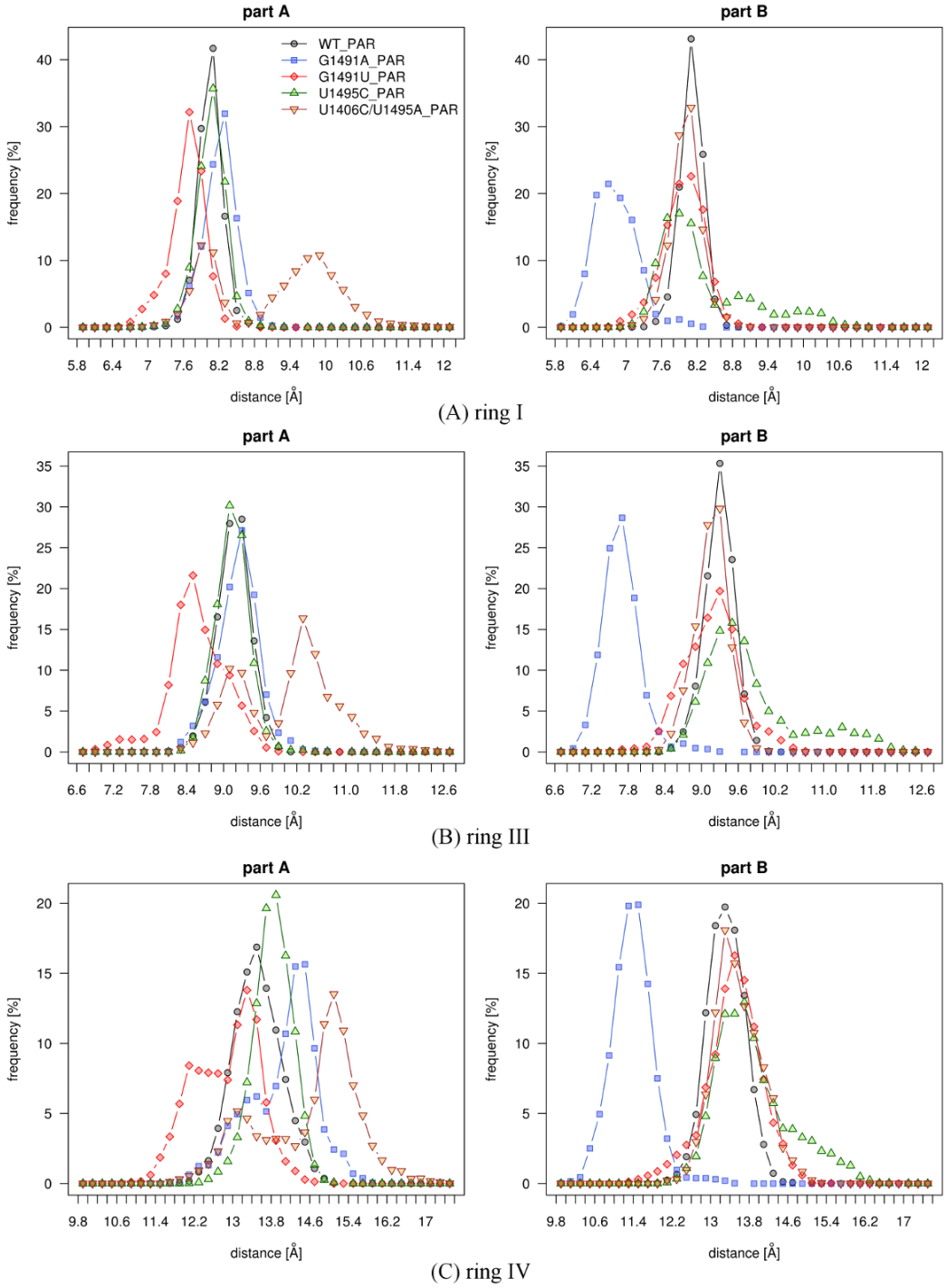


Figure 4.22: **Distance [Å] between the center of mass of A1492(P) and A1493(P) and the centers of mass of paromomycin's rings** — histograms of values measured in the simulations.

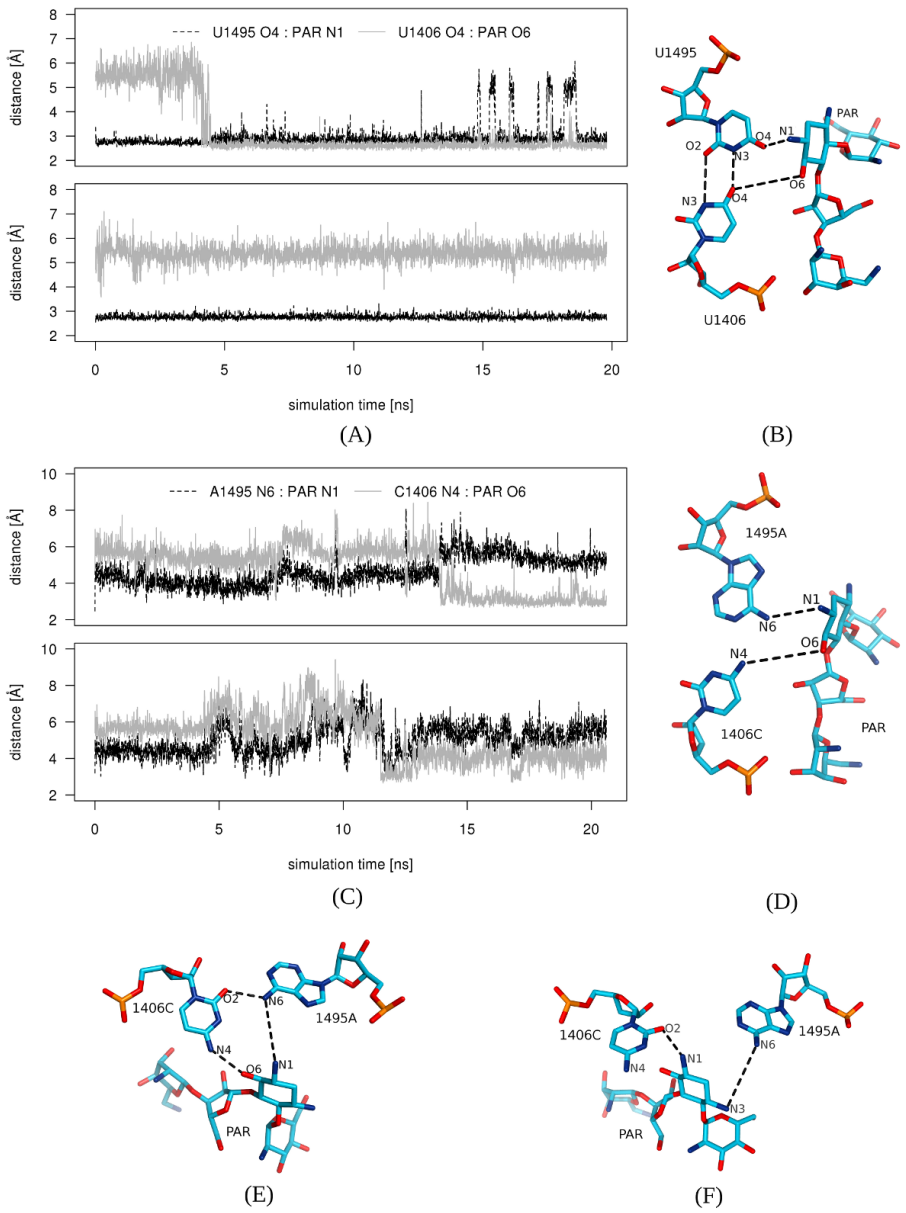


Figure 4.23: **Hydrogen bonds formed between ring II of paromomycin and RNA**. Distances between the atoms of the U1406·U1495 pair and paromomycin as a function of the simulation time along with the trajectory snapshots labeling the measured distances. Data taken from simulations *(A)–(B)* G1491A_PAR and *(C)–(F)* U1406C/U1495A_PAR. Two plots depict values for the two A-sites of the model. [Figure taken from our published results[211]]

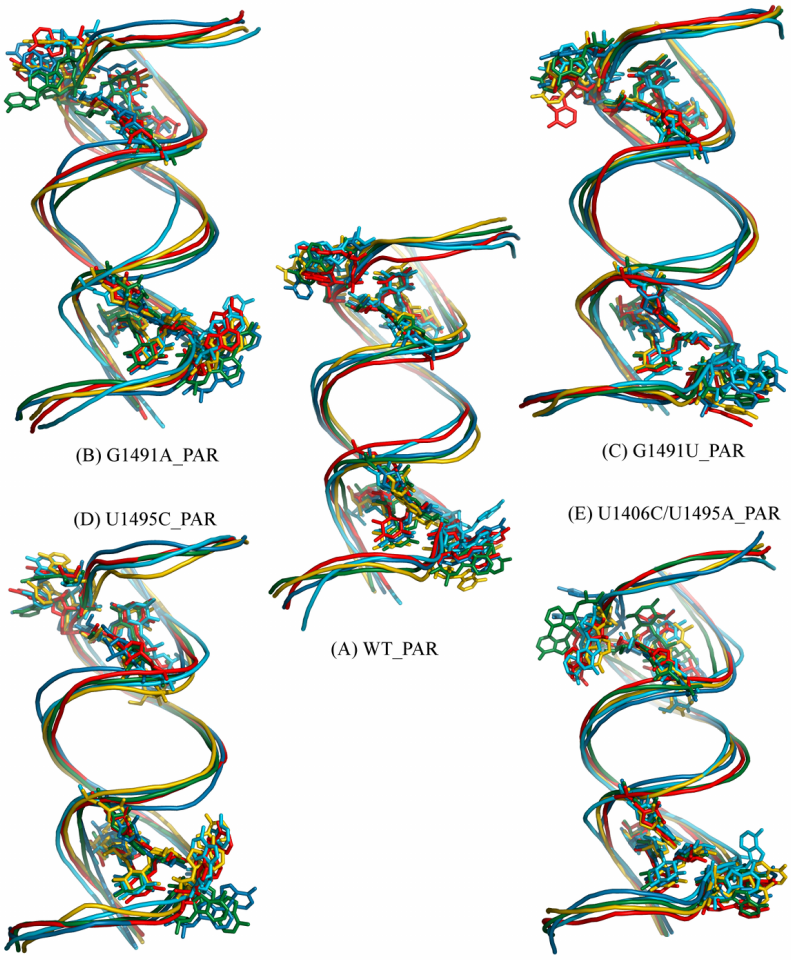


Figure 4.24: **Representative structures of clusters from the MD simulations of A-site:PAR complexes**, superimposed with regard to the phosphorous atoms. Structures from each cluster are colored differently: 1 — green, 2 — blue, 3 — red, 4 — yellow, 5 — cyan. Only paromomycin and bases A1492, A1493, and A1408 are shown in atomic detail. [Figure taken from our published results[211]]

Range of movement of A1492 and A1493 is increased in the complexes of mutated RNA A-site

These differences in stability of the complexes can be further strengthened by visualization of the clustering of conformations from the MD simulations (Figure 4.24). All simulations of the mutated structures allowed for a wider range of movement of A1492 and A1493 than WT_PAR. Especially in U1495C_PAR and U1406C/U1495A_PAR, the adenines were much closer to the flipped in state. It has been suggested that the effectiveness of aminoglycosides is correlated with the restriction of A1492 and A1493 to their flipped-out positions,[20,218] and therefore, the visible widening of the range of their movement in the simulations of mutated complexes versus the WT is in agreement with this hypothesis.

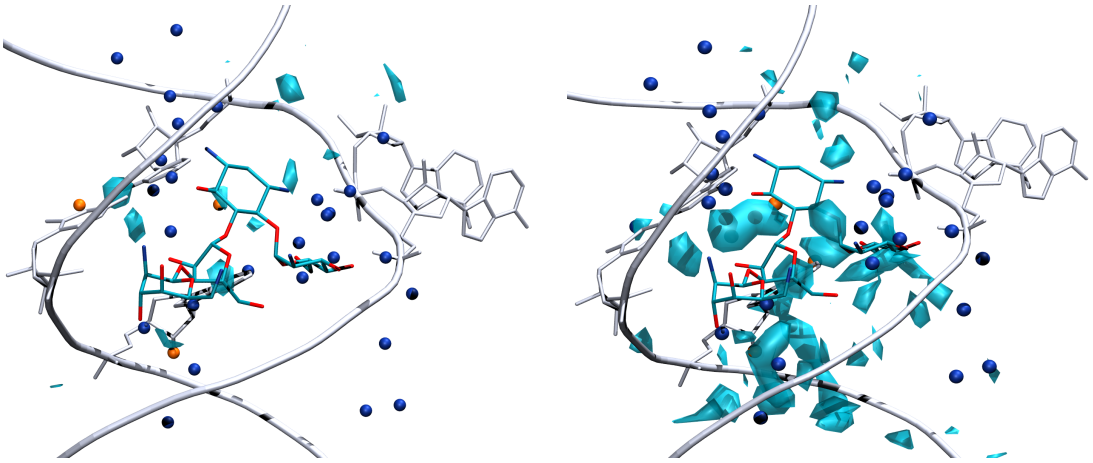
Experiments performed on the *eukaryotic* A-site with the A1491G substitution[15] demonstrated a minor decrease in the translation error frequency, while the presence of an aminoglycoside caused a 10-fold increase of the error rate. This suggests that a reverse mutation in bacteria would be very beneficial in terms of resistance and would carry a low cost of a slight increase in the translation error rate in the absence of the drug. Our results provide evidence for this theory. In the G1491A_PAR simulation, A1492 and A1493 moved closer

to the flipped-in state than in the wild type complex simulation (Figure 4.24), which is an indication of lower effectiveness of the drug.[20,218] Moreover, the flipped-in state was more occupied in the simulation without paromomycin (G1491A). This can potentially lead to a rejection of too many tRNA molecules and thus an increase in translation errors.

4.2.6 Water molecules inside the aminoglycoside binding site in RNA

The more resistance-causing mutation, the less water-mediated interactions are reproduced

In the complexes of aminoglycosides with the RNA A-site, non-direct (water-mediated) interactions are often critical.[31,142] There are seven such interactions reported in the wild type complex with paromomycin,[142] all formed with the neamine core (i.e., rings I and II). We calculated the average water density for the simulations of the A-site with different mutations, with and without paromomycin.



(A) G1491A (B) U1406C/U1495A

Figure 4.25: **Areas of high water density** (light blue, $\geqslant 0.23$ water molecule per Å^3) located in one part of the simulated structures, superposed on the first frame of the complexes with PAR; only the 1406, 1495, A1408, A1492, and A1493 bases are shown in atomic details; spheres show the positions of the crystal water oxygen atoms, those that were identified in the simulation are marked in orange (see Table 4.5). Hydrogen atoms were not shown for clarity of the image.
[Figure taken from our published results[211]]

The water dense areas in the G1491U, G1491A, and U1495C simulations were less numerous than in the U1406C/U1495A simulation (Figure 4.25). This might have been caused by the change in the shape of the binding site due to the shift of base pairing (as observed in G1491U and G1491A) or simply because A1492 and A1493 were positioned inside the cleft (in all singly mutated structures). In contrast, much more water dense areas were identified in the simulation of the structure with the double mutation (Figure 4.25B). Many of them were localized in the positions of paromomycin's atoms from the superimposed crystal structure. This suggests that the cleft was more "open" throughout the simulation and that the drug would have to expel many water molecules when binding to this doubly mutated A-site.

In the A-site:PAR crystal structure, 1J7T, there are many water molecules close to paromomycin. We checked whether our simulations of the complexes reproduced these molecules (Table 4.5). In the WT_PAR simulation, 12 areas of high water density were located in positions of the superimposed crystal water molecules, including all 7 that helped create the water-mediated polar contacts. Similar case was with the G1491U_PAR simulation, for

which only W19 was not reproduced. The other mutation of G1491 led to more changes in the interactions with paromomycin and only eight water dense areas were close to the positions of the crystal water molecules, of which only three represented the water-mediated hydrogen bonds. This shows that the G1491A mutation influenced the complex stability. Although in the U1406C/U1495A simulation, there were many water dense areas, in the simulation of the complex, only two water-mediated contacts were reproduced and only six crystal water molecules in total. This again demonstrates that the double mutation resulted in the weakest bound paromomycin.

Table 4.5: **Reproduction of the crystal water molecules in simulations of the A-site:PAR complexes.**

water molecule ↓	G1491U_PAR	G1491A_PAR	U1406C/U1495A_PAR	U1495C_PAR	WT_PAR
W1	+	−	+	+	+
W4	+	+	+	+	+
(W7)	(−)	(−)	(−)	(−)	(−)
W8 (W54) (*)	+(−)	−(+)	+(−)	−(−)	+(−)
W9 (W2) (*)	+(+)	−(+)	−(+)	+(+)	+(+)
W12	−	−	−	−	−
W13 (*)	+	−	−	−	+
W14 (*)	+	+	+	+	+
W15	−	−	−	−	−
W19 (*)	−	−	−	−	+
W20	−	+	−	+	+
W25	+	+	+	+	+
W27	+	−	−	−	−
W28	+	−	−	+	−
W32 (*)	+	+	−	+	+
W44 (*)	+	+	−	+	+
(W45)	(−)	(−)	(−)	(−)	(−)
W49	+	+	+*(near)*	+	+
W51	−	+	+	+	−

Selected water molecules of the 1J7T crystal structure and the corresponding high water density areas in the MD simulations with paromomycin. "+" denotes water density areas higher than 0.22 water molecules per Å^3 observed in the position of the corresponding crystallographic water molecule; "−" denotes lack of high water density in this position. Brackets denote water molecules located in the second symmetric part of the RNA fragment. [Table taken from our published results[211]]

(*) these water molecules mediate the non-direct hydrogen bonds between paromomycin and RNA in the original crystal structure, 1J7T[142]

4.2.7 Proposed modifications of paromomycin

Based on our observations from the MD simulations, some modifications to paromomycin can be proposed. Please refer to Figure 3.3A for numbering of paromomycin atoms, and to Table 4.4 for the hydrogen bonds formed in the WT A-site:PAR complex. Concerning the mutations of the U1406·U1495 bases, a substitution of the 2‴–NH_3^+ group with 2‴–$CH_2NH_3^+$ could improve the A-site:PAR stability, because it would allow the ring IV to interact with the phosphate groups of both G1405 and U1490 or G1491, and thus, paromomycin would anchor more in the mutated A-site. This modification would also increase the stability of the complexes with mutated G1491, since we observed that paromomycin ring IV in these complexes formed hydrogen bonds with the RNA backbone, i.e., interactions between 2‴–NH_3^+ of paromomycin and G1405(O2P), and/or 6‴–NH_3^+ and O2P of the mutated base 1491.

To strengthen the direct interactions with ring II, the 6–OH group of paromomycin could be substituted for CH_2OH. This would enable creation of a direct hydrogen bond between U1406(O4) and PAR(O6), instead of a water-mediated one observed in the WT structure[84,142] and in our WT_PAR simulation. Perhaps even farther strengthening could be achieved when swapping the amino group at 2′ position (ring I) with the OH substituent at 3′ position. The hydrogen bond between 3′–OH and A1492(O2P) was less stable in the G1491A_PAR and the U1406C/U1495A_PAR simulations than in the WT_PAR, and a 3′–NH_3^+ group would form a stronger interaction with the phosphate.

As has been shown, the least stably bound paromomycin was in the U1406C/U1495A_PAR simulation. One of the important interactions in the WT complex involve the 3–NH_3^+ of paromomycin (ring II) and the phosphate groups of either G1494 or A1493, which was not maintained in the simulation of the doubly mutated complex (after the equilibration, there was an increase in distance from ~ 3.9 Å up to 4.9 Å). Therefore, an extension at this position (i.e., 3–$CH_2NH_3^+$ instead of an amine group) could improve binding, which may also diminish the effect of the double mutation U1406C/U1495A.

4.2.8 Summary

The results from our MD simulations of the A-site model with different mutations showed that the mutations influence different features of the paromomycin binding site. We have demonstrated that a substitution at position 1491 changed the base pairing pattern, which led to a more compact binding site. The shape of this binding site appears to be an important factor of aminoglycoside affinity towards the A-site, and therefore the change introduced by the G1491 mutations hinders binding. The mobility of A1492 and A1493 was also affected in our G1491A and G1491U simulations, resulting in the adenines occupying the flipped-in state for most of the simulation time. Moreover, the G1491A mutation made the adenines move together more often than in WT. This observation helps explain the increased stop codon read-through of the bacterial ribosomes carrying this mutation — the recognition of a termination factor requires that only A1492 be flipped-out, while A1493 stays inside the helix. Since this conformational state of the adenines was sampled less often in the G1491A structure, it is more likely that a termination factor would be rejected by the mutated ribosome.

The outcomes of the simulations of structures with the mutated U1406·U1495 pair confirmed that these bases are important for aminoglycoside binding. They provide the negatively charged moieties that form hydrogen bonds with all aminoglycosides binding in this site. Therefore, when the uridines are mutated, an aminoglycoside may have difficulties in recognizing this cleft as its binding site. The double mutation introduced the most pronounced changes in our simulations, which is in agreement with the MIC studies that found that bacteria possessing this mutation were highly resistant to almost all aminoglycosides. Our U1406C/U1495A_PAR simulation showed that the complex with paromomycin was not stable and the range of movement of A1492 and A1493 was much wider than in the WT_PAR simulation. Aminoglycoside efficiency has been shown to correlate with the degree of adenine mobility restraint. Therefore, our simulations of the complexes qualitatively confirm this result.

4.3 Bacterial resistance mechanism 2: enzymatic aminoglycoside modification

To gain a more complete view on bacterial resistance mechanisms against aminoglycosides, we studied the most common mechanism, i.e., when bacteria produce enzymes (AMEs) that chemically modify aminoglycosides and inactivate them. We selected three enzymes: AAC(6′)-Ib, ANT(4′) and APH(3′,5″)-IIIa. The ANT representative is active as a homo dimer, and contains two aminoglycoside binding sites, both occupied by kanamycin A in our ANT_HOLO simulations. Conversely, AAC and APH are active in their monomeric states. Below, we present the results from these MD simulations of AME representatives and their complexes with kanamycin A and from a simulation of the A-site:KAN complex, which was conducted as a reference. For the detailed list of the selected structures and performed simulations, see Tables 3.2 and 3.3 in Chapter 3.

4.3.1 Conformations of bound kanamycin A

To study the conformational diversity of kanamycin in our simulations, we employed the same measure as in the case of paromomycin, i.e., we monitored the ϕ/ψ dihedral angles of the linkages (see Section 3.3.2 *Parameterization validation* for the definition). Figure 4.26 gathers the results from these measurements. Kanamycin's conformation in the complex with the rRNA A-site was similar to the one that we found to be the most occupied in water (compare Figure 4.26, top, with Figure 3.4C–D). Linkage I/II acquired the same ϕ/ψ values for both aminoglycosides complexed with the native A-site (KAN and PAR; compare Figure 4.26, top with Figure 4.12, top). Notably, the kanamycin's conformation in two of the studied enzymes (AAC and APH) was very similar to the one in the A-site. However, binding to ANT required a drastic conformational change in linkage I/II and also made linkage II/III more dynamic. This suggests that ANT binds kanamcyin A less tightly than other enzymes. Previously reported X-ray and NMR experiments[67,68,79] showed that the conformation of several aminoglycosides bound to some AMEs is significantly different from the one adopted in the A-site. This fact was utilized to propose some conformationally constrained aminoglycoside derivatives[80,219] that disabled their binding to AMEs. However, that strategy is very limited, since even if an aminoglycoside changed its conformation when binding to one AME, in the complexes with other AMEs the conformation would be the same as acquired inside the A-site, and therefore constraining aminoglycosides would not prevent binding to all possible bacterial AMEs.

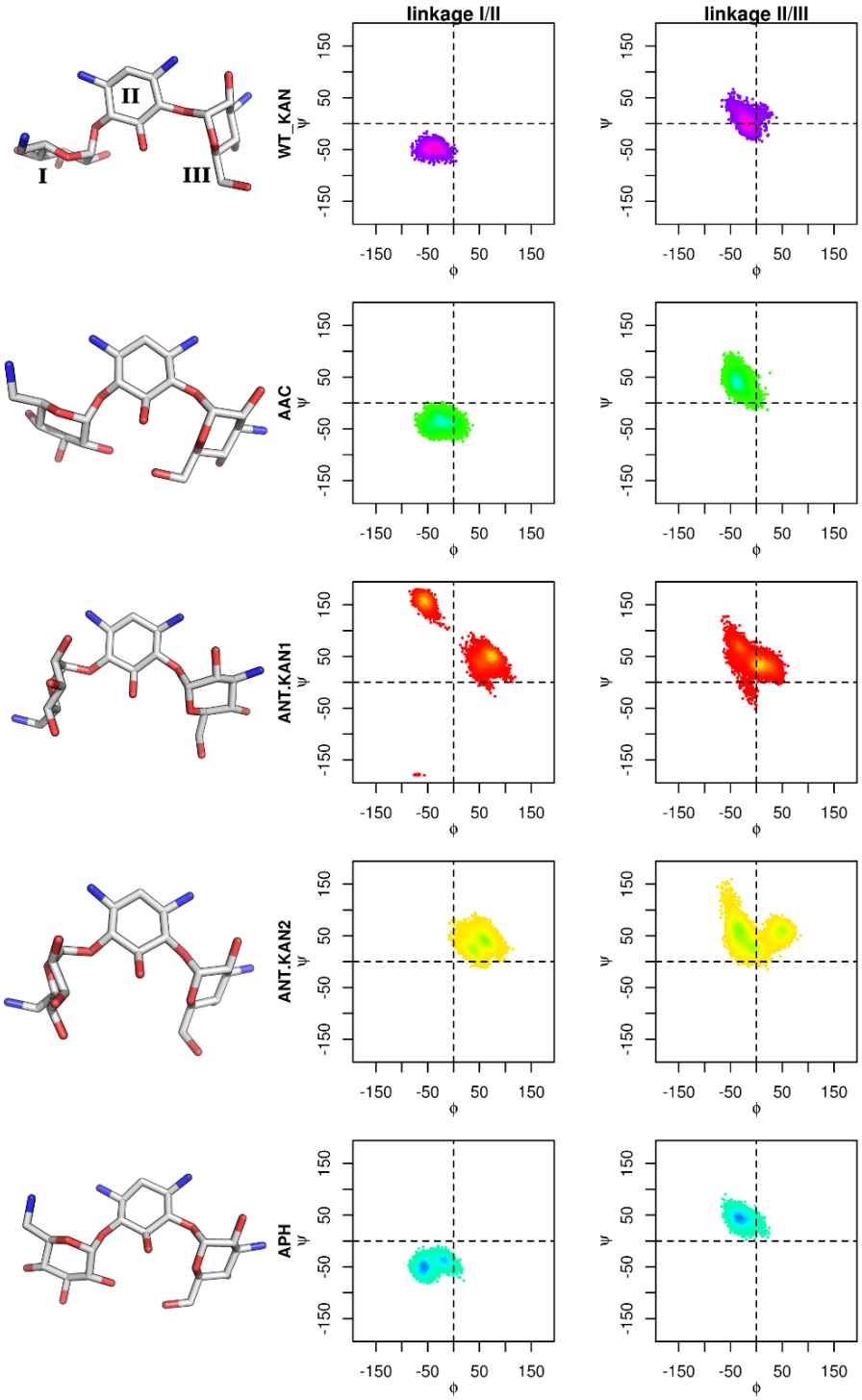


Figure 4.26: **Conformations of kanamycin A bound to different receptors.** On the left, the most populated conformations from the simulations are presented. The graphs depict changes of the dihedral angles of the linkages between the rings.

4.3.2 Internal dynamics of enzymes

Flexibility patterns differ among AMEs

The rRNA A-site bulge is flexible, and we have shown that even one-base substitutions can change its internal dynamics. To investigate the dynamic features of AMEs, we looked at the differences in RMSF between the APO (i.e., binary complex: enzyme + cofactor) and HOLO (i.e., ternary complex: enzyme + cofactor + kanamycin A) states of the enzymes. This difference in the residue fluctuations of each enzyme is presented in Figure 4.27. The monomeric enzymes, i.e., AAC and APH, displayed the biggest changes in the residue mobility upon kanamycin A binding, especially the loops located over the binding site were restrained in the HOLO states. On the contrary, the changes in RMSF between the APO and HOLO simulations were much smaller for the dimeric ANT. Moreover, this enzyme was the least dynamic from all three structures, as indicated by the range of RMSF values presented in Figure 4.28.

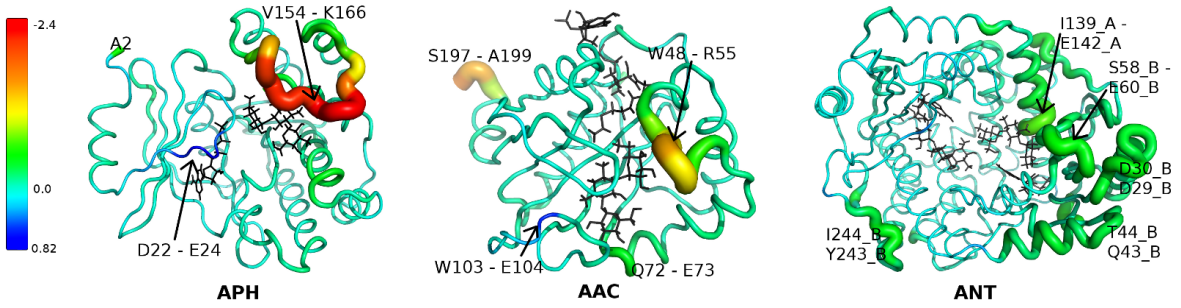


Figure 4.27: **Difference in internal dynamics between the HOLO and APO states of the enzymes**, presented as the difference in per-residue RMSF [Å] calculated for all simulations; the negative values point to more mobile residues in the APO state. The residues that had abs(ΔRMSF) $> 0.5, 0.3$ or 0.4 Å for APH, AAC and ANT, respectively, are labeled.

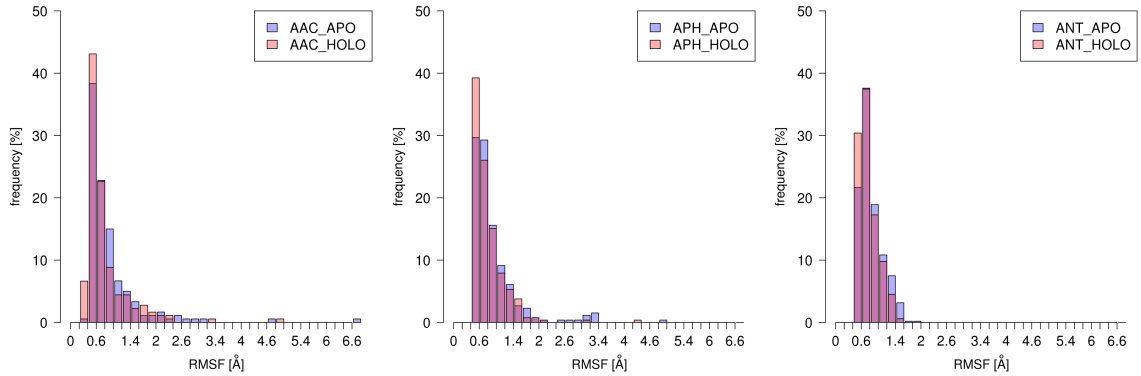


Figure 4.28: **Histograms of the per-residue RMSF values from the simulations of AMEs**; comparison between the residue mobility in the APO and HOLO states.

We studied the enzymes' dynamics with our new tool, GeoStaS, that analyzes the differences and similarities in molecular motions. The GeoStaS analysis of the trajectories shows the division of AMEs into dynamic domains (presented in Figure 4.29). The results indicate that the flexibility patterns did not change upon kanamcyin binding for both ANT and AAC. In contrast, the APH enzyme displayed significantly different pattern of fluctuations in the APO and HOLO simulations. To analyze this finding further, we collected all other experimental structures of AAC and APH, that were available in PDB. Then we performed PCA

on those structures and projected the conformations from the simulations onto the identified principal components. In the case of APH X-ray structures, both PC 1 and PC 2 describe the movement of the aminoglycoside binding loop (i.e., residues from around Asp 150 to Asp 162) and of the cofactor binding loop (i.e., residues from around Asp 22 to Ser 27), as illustrated in Figure 4.30A. In the case of AAC, PC1 shows large movements within the residues positioned close to the cofactor binding site and the loop over the aminoglycoside binding site (i.e., residues from Glu 47 to Arg 55). Indeed, three residues from this loop are not present in several X-ray structures. PC2 of the AAC conformations describes a rather moderate movement of the loop over the aminoglycoside binding cleft. The first mode is depicted in Figure 4.30B. The conformations from X-ray studies and simulations projected onto PC1 *versus* PC2 plane yield the so-called conformer plots, presented in Figure 4.31. The results demonstrate that the range of conformational sampling of APH and AAC was diminished upon binding of kanamycin A. This restriction in dynamics was much more pronounced in the case of APH. Moreover, the sampled conformations were close to those that were resolved in the complexes with some other ligands, which demonstrates the high flexibility of APH and AAC.

Only one experimental structure has been resolved of the dimeric ANT. Consequently, the conformer plot analysis were not suiting. Thus, we performed clustering of all simulated conformers sampled along the six independent simulations to investigate whether some

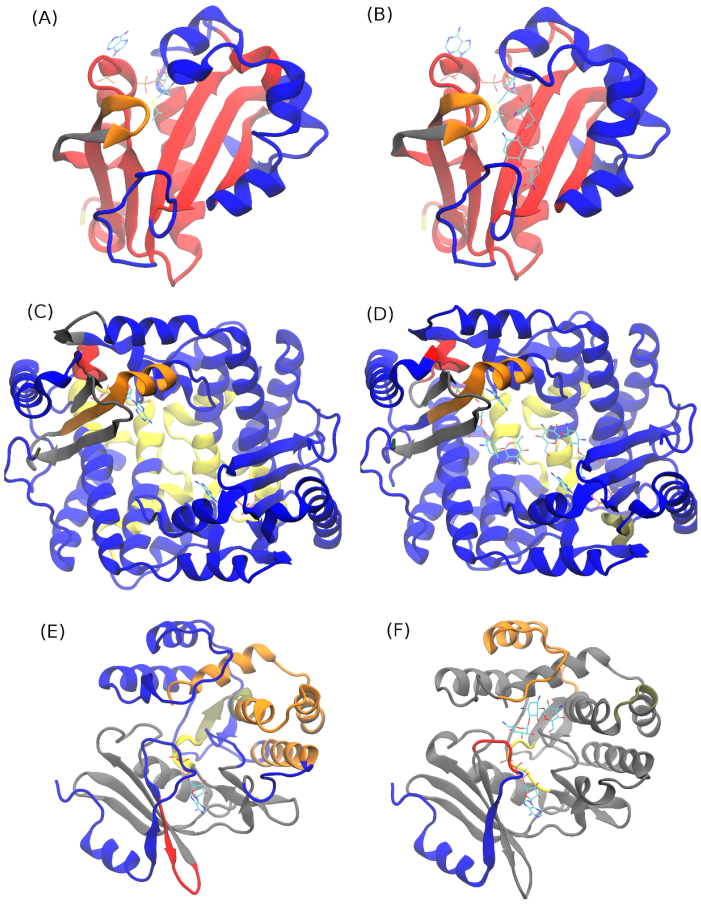


Figure 4.29: **Dynamic domains found by GeoStaS based on conformations from the AME trajectories**: *(A)* AAC_APO, *(B)* AAC_HOLO, *(C)* ANT_APO, *(D)* ANT_HOLO, *(E)* APH_APO, and *(F)* APH_HOLO. Each structure is colored independently. The enzymes are presented in cartoon, and the ligands — in sticks.

conformations appear both in the APO and HOLO simulations. Figure 4.32 shows that ANT sampled different states in the simulations with and without kanamcyin A, i.e., clusters were not shared between the APO and HOLO simulations.

Our results demonstrate that each of the simulated enzymes displayed different flexibility pattern, which perhaps determines the way of gaining specificity for aminoglycosides. The studied dimer, ANT, was the least flexible, which could explain the unique order of binding of the substrates: ATP binds after an aminoglycoside. The binding of a drug to the binary complex ANT:ATP would probably be more difficult since there could be not enough space for the drug to adapt to the non-flexible binding site. Another hypothesis has been recently formulated, as a result of experiments and short simulations of the binary and ternary complexes of ANT(4′).[220] This new hypothesis suggests that binding of an aminoglycoside is two-fold: first, when there is no ATP, an aminoglycoside binds non-specifically, and the relocation to the main binding site is triggered by the presence of the nucleotide.

In our simulations, the monomeric AAC was more dynamic than ANT. Specifically the loops positioned over the aminoglycoside binding site were flexible. However, our results show that this pattern of internal dynamics did not change when kanamycin A bound to AAC. Thus in this case, the global conformation of the enzyme could be similar when forming complexes with different drugs. Moreover, since the structure of the studied AAC resembles the structures of other enzymes from the AAC(6′) subfamily, this mobility pattern may be similar for the subfamily. However, this should be investigated further.

The APH enzyme was the most flexible in the APO simulations, and it adopted many conformations that resembled the experimental conformations of its complexes with different aminoglycosides. While the NMR experiments suggested a highly disordered structure of APH,[75] the X-ray studies were able to capture well the unliganded enzyme and enzyme:cofactor complexes. Our results may explain this contradiction — APH is in a dynamic equilibrium between different thermally accessible conformations thus one may be able to observe a well defined conformation in a crystal. Our analysis demonstrated that when complexating with a drug, APH becomes more rigid, which indicates that this process carries a high entropic cost. We hypothesize that it is probably in order to lower this entropic cost that APH(3′,5″) sometimes forms dimers.[73]

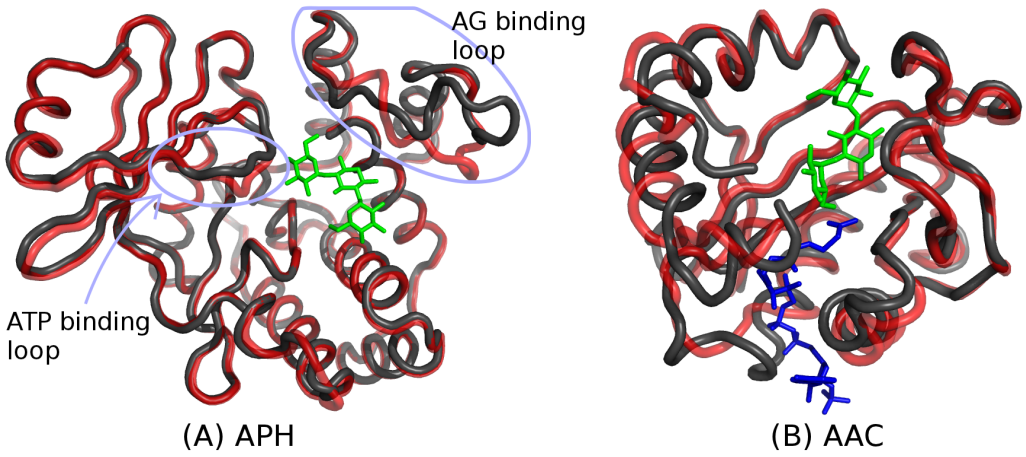


Figure 4.30: **Visualization of the first PC for the APH and AAC X-ray structures**, presented as the backbone trace of two most extreme conformations, colored black and red, respectively. The most dynamic loops in APH are encircled and labeled; kanamycin A (green sticks, in both structures) and ACO (blue sticks, in AAC) are superimposed for reference.

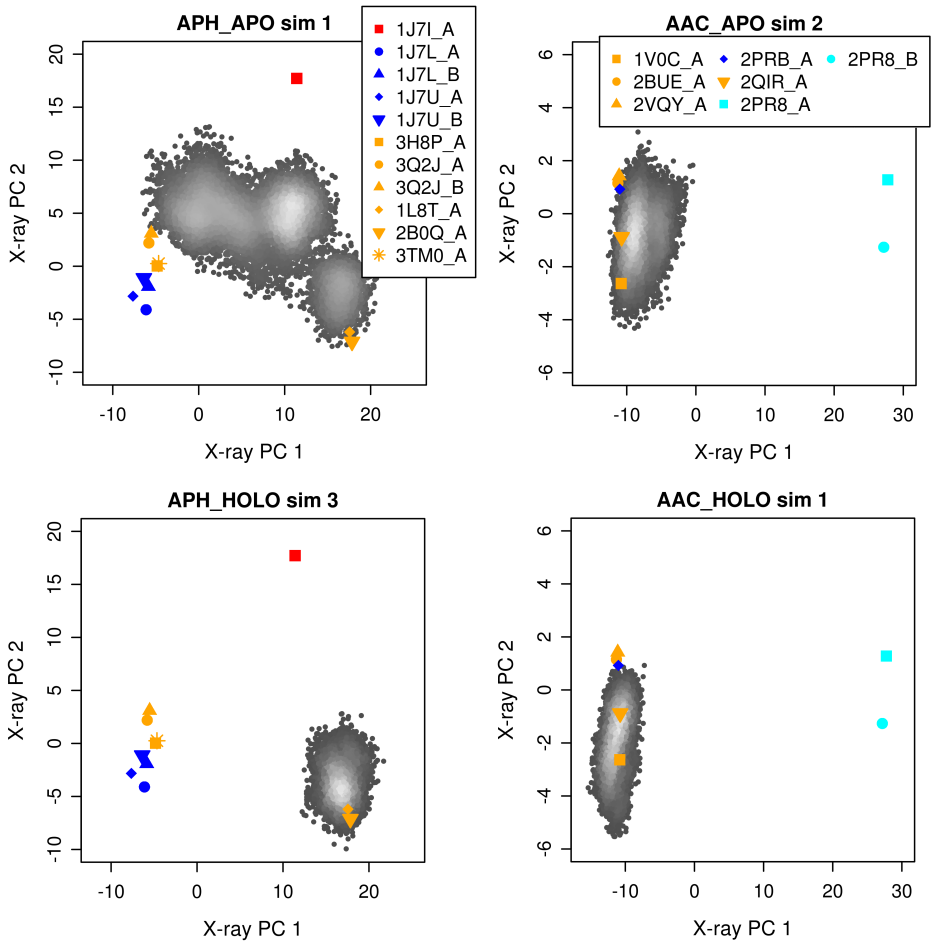


Figure 4.31: **Comparison of the trajectory-derived conformations (shaded gray dots) with the experimental structures of the AAC and APH enzymes (colored points)**. The plots show differences in conformations based on the two first PCs, which constitute 74.2% and 98.7% of the conformation variance for APH and AAC X-ray structures, respectively. The structure of the unliganded wild type enzyme is marked by a red square, the enzymes complexed with a coenzyme only are colored blue, the ternary complexes (enzyme + coenzyme + antibiotic or inhibitor) are in orange, and the unliganded mutated enzyme in cyan.

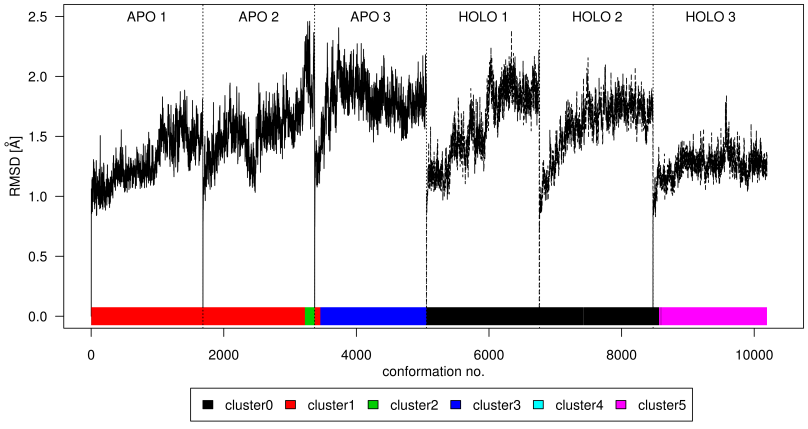


Figure 4.32: **Clustering of ANT conformations sampled in all the simulations** — each color bar in the bottom of the graph represents a different cluster. Black line shows the RMSD for each of the trajectories, with respect to each starting conformation.

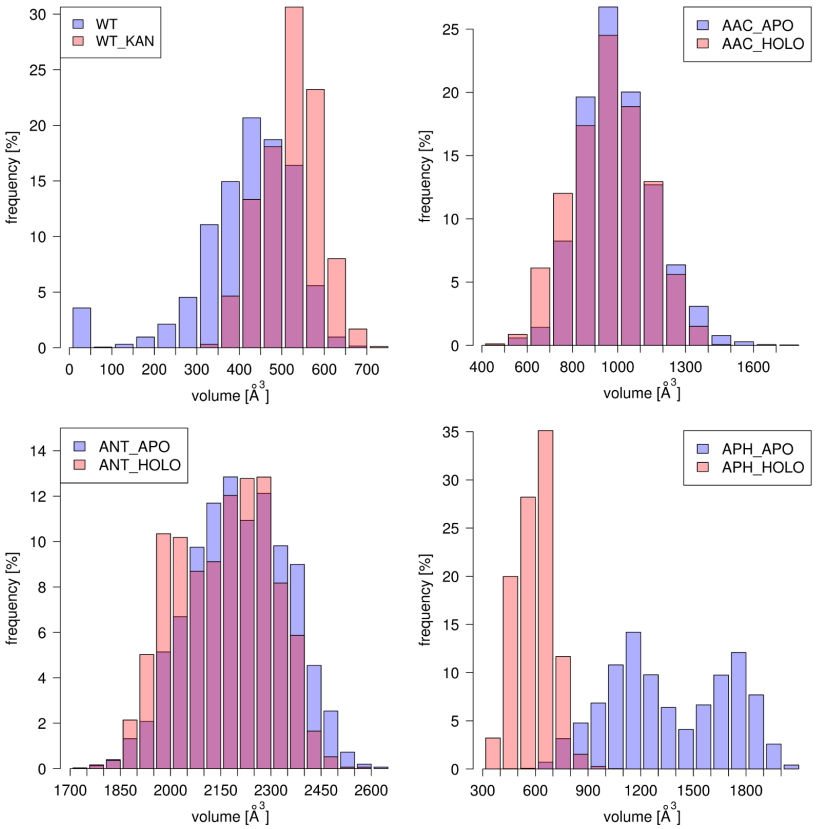


Figure 4.33: **Histograms of aminoglycoside binding site volume [Å^3] in the simulations of the A-site and AMEs**.

4.3.3 Shapes and sizes of aminoglycoside binding pockets in AMEs

Volume of binding pocket is greatly reduced upon aminoglycoside binding in APH

We measured the volume of kanamycin A binding sites to compare the pockets among the selected AMEs and between AMEs and the rRNA A-site. Therefore, when defining the space of the pocket on the RNA, we have decided to include only the space that would be occupied by kanamycin A. However, we note that an aminoglycoside of different shape and size could also fit into the pocket and could give different volume values because this RNA structure is flexible and the cavity is long and wide. Figure 4.33 compares the volumes of the binding pockets in AMEs from the APO and HOLO simulations of each enzyme, and the volume of kanamcyin A binding site from the simulations of the A-site. In the RNA:KAN complex, the histogram was narrowed to the larger values compared to the simulation of the bare RNA, because the adenines A1492 and A1493 were restricted to their flipped-out conformation. Both for AAC and ANT the cleft's volume did not change upon binding of kanamcyin. In contrast, the space available for an aminoglycoside in APH_HOLO was much smaller than in APH_APO. Together with the previously described results on the internal dynamics of this enzyme, this suggests that the residues around the aminoglycoside binding site in APH are very flexible and the enzyme adapts to the shape of aminoglycosides when forming the complex, which suggests an induced fit mechanism of binding.

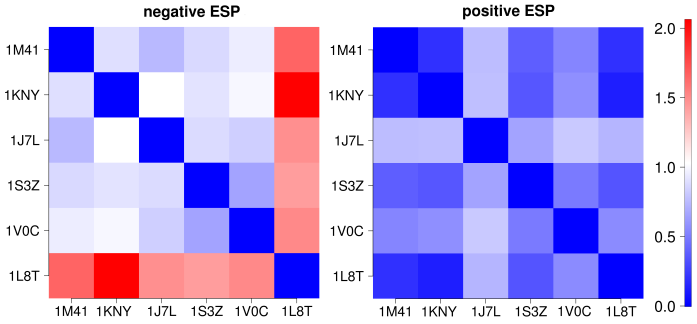


Figure 4.34: **Comparison of the electrostatic potential in the aminoglycoside binding clefts of different AME X-ray structures** (indicated by PDB ID, described in Table 3.2). The scale represents the root mean square distance between the calculated potentials, relative to rings I and II of kanamycin A (i.e., the neamine core).

4.3.4 Electrostatic potential inside aminoglycoside binding sites

The binding of aminoglycosides to the A-site has previously been shown to be dominated by electrostatic interactions,[217] due to the high charges on both, aminoglycosides and RNA. Therefore, we analyzed the electrostatic potential (ESP) inside the binding sites of the selected biomolecules. In the case of AMEs, the ESP was calculated directly for each of the static X-ray structures (see Table 3.2 in Chapter 3), through the Poisson-Boltzmann continuum electrostatics methodology (as described in Chapter 2, Section 2.7.1 *Poisson-Boltzmann model*). We then used the Surface Diver[179] software to compare the calculated ESP from the point of view of the neamine core positioned inside the aminoglycoside binding site in AMEs (see Section 3.4 *Data analysis* for the description of this method). Neamine is the common part of the majority of aminoglycosides, therefore each enzyme should create a very similar electrostratic environment for this moiety.

Figure 4.34 shows the root-mean-square (RMS) distance between the ESP values for the pairs of enzymes, separately for the part that originates from the positive and negative charges. The maximal difference for the negative part of ESP was much larger (RMS distance $\leqslant$ 2.06) than for the positive part (RMS distance $\leqslant$ 0.805). This is because the binding sites need to be negatively charged to attract the positively charged aminoglycosides, and therefore the positive part of the ESP was negligible. In general, the enzymatic binding sites differed in the negative part of ESP, with the 1L8T structure being the most distinct from the others. It even differed from 1J7L, which is a dimeric version of the same enzyme. This might be caused by the loop positioned over the binding site, which is "closed" in the monomeric and "open" in the dimeric X-ray structure.

Table 4.6: **Common chemical features of the amino acids inside the aminoglycoside binding sites in AMEs.**

	number of chemical groups				
chemical group type ↓	1KNY	1M4I	1S3Z	1V0C	1L8T
acceptor	15	14	12	10	16
donor	2	6	13	9	11
aliphatic	1	4	5	0	3
aromatic	0	3	4	6	3

The number of chemical groups of each type is shown based on the classification of amino acids presented in Ref.[221] The enzymes are labeled by their PDB code — for description see Table 3.2.

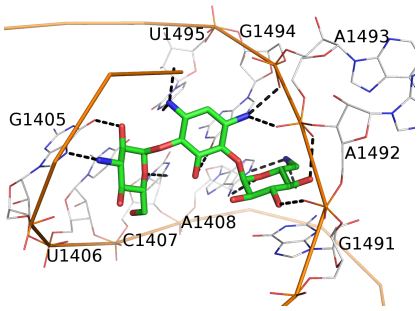
4.3.5 Short-range interactions formed by the receptors with kanamycin A

Comparison of AME:AG X-ray complexes

Next, we investigated how the different AMEs interact with aminoglycosides. We compared the amino acids that interact with aminoglycosides in different X-ray structures of the AME:AG complexes. While the amino acids that are within 5 Å from the bound antibiotic differed in type, the chemical groups they display were similar. As presented in Table 4.6, the most common groups were acceptors of hydrogen bonds. This is due to aminoglycosides having many NH_3^+ groups, which can act as donors. In the three structures, 1S3Z, 1V0C, and 1L8T, apart from hydrogen bond acceptors there was also a similar number of donors. This can possibly help these enzymes be more promiscuous. The other types of chemical groups were poorly represented inside the AME binding sites. Notably, even aromatic groups were rare, although they could form stacking interactions with the rings of aminoglycosides. Only in the binding site of 1V0C (AAC(6′)-Ib) a substantial number of aromatic groups could be found.

Kanamycin A in the rRNA binding site

The performed simulations (see Table 3.3) focused on interactions of AMEs with one aminoglycoside, kanamycin A. However, first, we studied the A-site:KAN complex, to be able to compare and relate it to the AME complexes. Although kanamycin A has a slightly different chemical structure and shape than paromomycin, it binds to the rRNA A-site in a similar manner, utilizing mainly its two first rings (eight out of eleven hydrogen bonds are formed with rings I and II).[31] In our WT_KAN simulation, all the hydrogen bonds from the crystal structure were maintained, even though some lasted shorter in one of the binding sites of the model than in the other, as can be seen from Table 4.7.



atoms involved in H-bond	occupancy of H-bond [% of simulation time]		X-ray distance [Å]	
KAN atom ... RNA atom	part A	part B	part A	part B
O2″...G1405(O6)	66	95	2.78	3.09
N3″...G1405(N7)	72	74	3.06	2.91
N6′...A1408(N1)	72	36	2.89	3.53
O3′...A1492(O2P)	91	93	2.63	2.94
O4′...A1492(O2P)	1	55	3.89	4.65
N3...A1493(O1P)	21	92	3.07	3.04
N6′...A1493(O2P)	0	53	4.72	3.65
O4′...A1493(O2P)	91	33	2.65	2.7
N3...A1493(O2P)	68	2	4.46	3.65
N3...G1494(O2P)	91	90	2.89	2.77
N3...G1494(N7)	73	48	2.81	2.66
N1...U1495(O4)	83	97	2.57	2.63

Table 4.7: **Direct hydrogen bonds formed between kanamycin A and the ribosomal RNA A-site**. Only the interactions between kanamycin and the receptor's residue that lasted at least 50% of the simulation time in at least one of symmetric parts of the receptor are shown. The location of the bases relative to kanamycin A is illustrated in the left panel.

AMEs mimic hydrogen bonding with KAN from the A-site:KAN complex

Next, we analyzed the simulations of the complexes that AMEs formed with kanamycin A. Compared to the complex formed with the A-site, slightly less hydrogen bonds were formed in the AAC_HOLO simulations (nine or eight, depending on the simulation run), while in the ANT_HOLO simulations, they varied from 6 to 12. The APH:KAN complex was the most stable, with 14 hydrogen bonds in each of the APH_HOLO simulation runs. The complete lists of hydrogen bonds formed in the binding sites are presented in Tables 4.8–4.10.

Naturally, since the aminoglycosides are positively charged in physiological pH, they are attracted by the negatively charged RNA backbone. We found that AMEs mimic these interactions quite well (Figure 4.35A–B). During the simulations, the enzymes formed many hydrogen bonds utilizing the Asp and Glu side chains. However, there were several cases where other amino acids formed equally stable interactions (Figure 4.35A), e.g., Ser 94 in the ANT_HOLO simulation (in both monomers), Ser 98 in AAC_HOLO and Phe 264 in APH_HOLO. The first two mentioned interactions were created through carbonyl groups of the backbone. Phe 264 is the terminal residue in APH(3′,5″), and therefore, it utilized its backbone carboxyl group, forming hydrogen bonds through either of the oxygens. This interaction was shown to be conserved within the APH(3′) subfamily.[222]

Table 4.8: **Direct hydrogen bonds formed between kanamycin A and AAC(6′)-Ib.** Only the interactions between kanamycin and a receptor's residue that lasted at least 50% of the simulation time in at least one of the simulations are shown. See Figure 4.35 for the kanamycin A atom numbering and the location of the amino acids.

atoms involved in H-bond	**occupancy of H-bond** [% of simulation time]		
KAN atom . . . enzyme atom	**sim 1**	**sim 2**	**sim 3**
O2″. . .Glu73(OE1)	0	25	0
O2″. . .Glu73(OE2)	0	33	0
O4″. . .Glu73(OE1)	15	3	39
O4″. . .Glu73(OE2)	29	0	26
N3″. . .Glu73(OE1)	29	60	26
N3″. . .Glu73(OE2)	24	48	13
O5. . .Gln91(NE2)	91	96	95
N1. . .Ser98(O)	0	57	17
N6′. . .Asp115(OD2)	96	92	95
N3. . .Asp115(OD2)	95	86	88
N3. . .Asp152(OD1)	55	19	58
N3. . .Asp152(OD2)	75	19	62
O4′. . .Asp179(OD1)	31	23	3
O4′. . .Asp179(OD2)	20	21	3
N6′. . .ACO(O)	87	85	83

Table 4.9: **Direct hydrogen bonds formed between kanamycin A and APH(3′,5″)-IIIa**. Only the interactions between kanamycin and a receptor's residue that lasted at least 50% of the simulation time in at least one of the simulations are shown. See Figure 4.35 for the kanamycin A atom numbering and the location of the amino acids.

atoms involved in H-bond	**occupancy of H-bond** [% of simulation time]		
KAN atom . . . enzyme atom	**sim 1**	**sim 2**	**sim 3**
N6′. . .Glu157(OE1)	66	31	42
N6′. . .Glu157(OE2)	70	70	29
N3. . .Glu157(OE1)	15	0	66
N3. . .Glu157(OE2)	86	0	33
O5. . .Glu157(OE1)	0	57	0
O5. . .Glu157(OE2)	0	33	0
N1. . .Asn158(O)	92	36	94
N1. . .Glu160(OE2)	100	96	100
N3″. . .Glu160(OE1)	90	95	97
O2″. . .Glu160(OE2)	100	100	100
O3′. . .Asp190(OD1)	0	40	95
O3′. . .Asp190(OD2)	0	40	3
N3″. . .Arg226(O)	41	0	82
N3″. . .Glu230(OE1)	51	57	61
N3″. . .Glu230(OE2)	44	65	42
O4″. . .Glu230(OE1)	26	52	0
O4″. . .Glu230(OE2)	18	44	0
N3. . .Asp261(O)	28	85	72
N1. . .Glu262(OE2)	100	92	100
N3. . .Glu262(O)	64	0	15
N6′. . .Phe264(O)	0	67	63
N6′. . .Phe264(OXT)	0	78	82
N3. . .Phe264(O)	67	76	68
O4′. . .Phe264(O)	50	0	1
O4′. . .Phe264(OXT)	52	0	0
O3′. . .ATP(O3G)	99	0	3
O4′. . .ATP(O3G)	0	0	90

Table 4.10: **Direct hydrogen bonds formed between kanamycin A and ANT(4′)-I.** Only the interactions between kanamycin and a receptor's residue that lasted at least 50% of the simulation time in at least one of the simulations are shown. The values in brackets refer to the hydrogen bonds formed in the second monomer. See Figure 4.35 for the kanamycin A atom numbering and the location of the amino acids.

atoms involved in H-bond	**occupancy of H-bond** [% of simulation time]		
KAN atom . . . enzyme atom	**sim 1**	**sim 2**	**sim 3**
O4′. . .Glu52(OE2)	0 (0)	86 (0)	0 (0)
N6′. . .Glu52(OE2)	0 (0)	0 (0)	99 (0)
O2′. . .Glu67(OE1)	50 (88)	100 (62)	7 (28)
O2′. . .Glu67(OE2)	50 (3)	0 (31)	1 (73)
N3. . .Glu67(OE1)	66 (46)	99 (56)	27 (59)
N3. . .Glu67(OE2)	66 (72)	14 (48)	86 (59)
O2′. . .Lys74(NZ)	0 (43)	0 (6)	0 (70)
N6′. . .Glu76(OE1)	61 (12)	6 (1)	92 (56)
N6′. . .Glu76(OE2)	56 (96)	11 (99)	54 (73)
O3′. . .Glu76(OE2)	0 (0)	76 (0)	0 (0)
N3. . .Glu76(OE2)	0 (56)	0 (72)	0 (0)
N1. . .Glu93(OE1)	13 (0)	0 (0)	88 (0)
N1. . .Glu93(OE2)	29 (0)	0 (0)	6 (0)
N3″. . .Ser94(O)	0 (0)	0 (0)	89 (1)
N3″. . .Asp95(OD1)	8 (0)	0 (39)	64 (4)
N3″. . .Asp95(OD2)	2 (3)	0 (24)	38 (0)
O2′. . .Glu141(O)	0 (0)	0 (0)	88 (0)
O3′. . .Glu141(O)	66 (0)	0 (0)	0 (0)
N3. . .Glu141(OE1)	27 (0)	0 (0)	0 (60)
N3. . .Glu141(OE2)	35 (43)	99 (2)	99 (37)
O2″. . .Glu142(OE1)	58 (71)	72 (12)	98 (92)
O2″. . .Glu142(OE2)	31 (22)	41 (36)	1 (13)
N1. . .Glu142(OE1)	0 (11)	0 (21)	0 (3)
N1. . .Glu142(OE2)	9 (33)	1 (12)	94 (80)
N3″. . .Glu142(OE1)	0 (0)	0 (42)	0 (0)
N3″. . .Glu142(OE2)	0 (0)	0 (24)	0 (0)
O3′. . .Glu145(OE1)	1 (30)	1 (52)	0 (15)
O3′. . .Glu145(OE2)	1 (52)	21 (76)	93 (9)

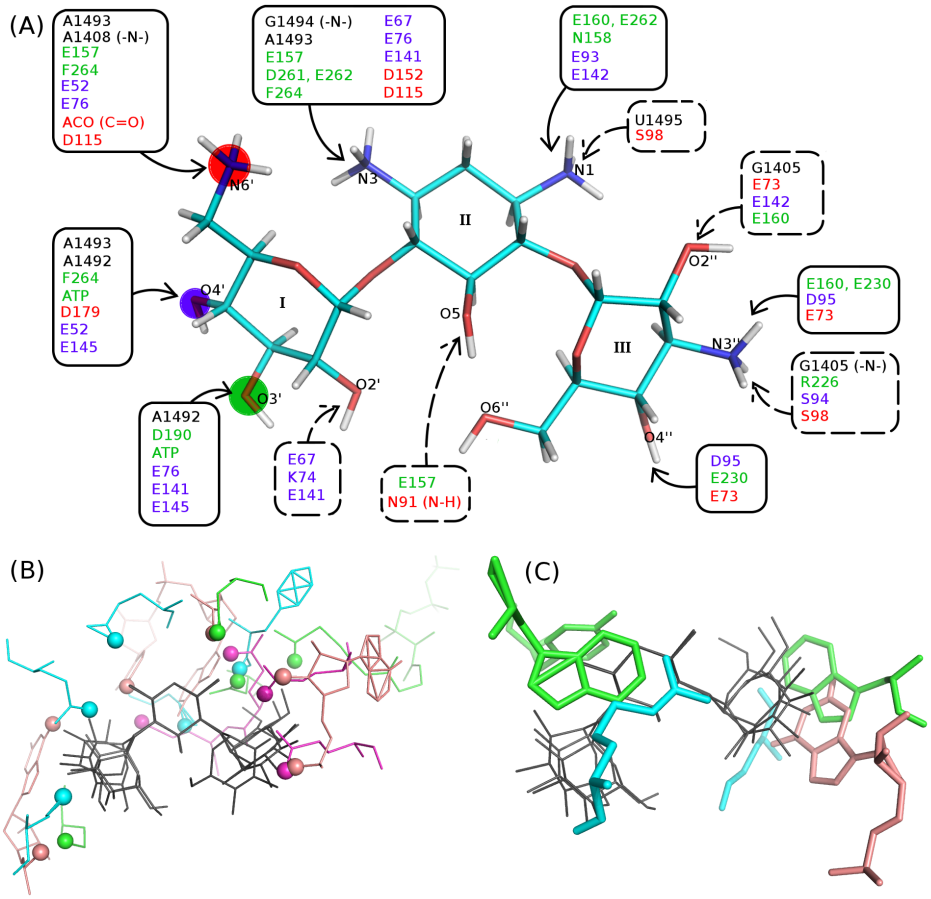


Figure 4.35: **Similar interactions formed with kanamycin A in different binding sites**.
(A) The hydrogen bonds lasting at least 50% of the simulation time formed with kanamycin A in different binding sites: rRNA A-site — black; AAC(6′)-Ib — red; APH(3′,5″)-IIIa — green; ANT(4′) — blue. Although the amino acids involved in forming the hydrogen bonds differ, the types of interactions were similar: the solid lines encircle the interactions with the phosphate or carboxyl groups, and the dashed lines — with carbonyl group. Few exceptions are denoted in parentheses next to the residue name. The colored circles around kanamycin A atoms mark the modification sites for each of the AMEs.
(B) The oxygen atoms that formed the hydrogen bonds (small spheres) during the simulations are positioned similarly in space, as illustrated by the average structures of receptors superimposed with respect to ring II of kanamycin A (dark gray lines): RNA (pink), AAC (green), ANT (magenta) and APH (cyan).
(C) The non-polar interactions in various receptors were formed with residues that have similar location, relative to kanamycin A (coloring and representation is the same as in panel *B*).

Non-polar interactions with KAN are similar in different receptors

Another type of interaction formed between RNA and aminoglycosides is stacking.[31,142] While classic stacking occurs between two aromatic moieties, sugar–aromatic stacking is also frequently observed.[223] In the A-site complexed with an aminoglycoside, ring I fits in the vacant space next to the flipped-out A1492 and A1493, and it stacks on top of G1491 (Figure 4.36A). Since this interaction is common to all aminoglycosides that bind to the A-site, we have checked whether it is copied by the enzymes (see Section 2.6 *Intra-molecular interactions* in Chapter 2 for the measurement details). In the WT_KAN simulation, the stacking between ring I and G1491 below was maintained throughout the simulation in both parts of the structure (mean distance: 4.15 ± 0.21 Å, mean cosine: 0.977 ± 0.021). In the AME simulations, only in the AAC binding site the aromatic residues were close enough to kanamycin to form stacking. These were: Trp 49 (stacked with ring I of the drug), Tyr 93 and Trp 102 (both stacked with ring II) (Figure 4.36B). These interactions were generally maintained throughout the simulation, with some short events of Trp 102 receding from ring II of kanamycin (Figure 4.36C). In the other studied enzymes no stacking was observed, although in APH_HOLO, van der Waals interactions were formed with the drug through two arginines, Arg 211 and Arg 226 (colored cyan in Figure 4.35C).

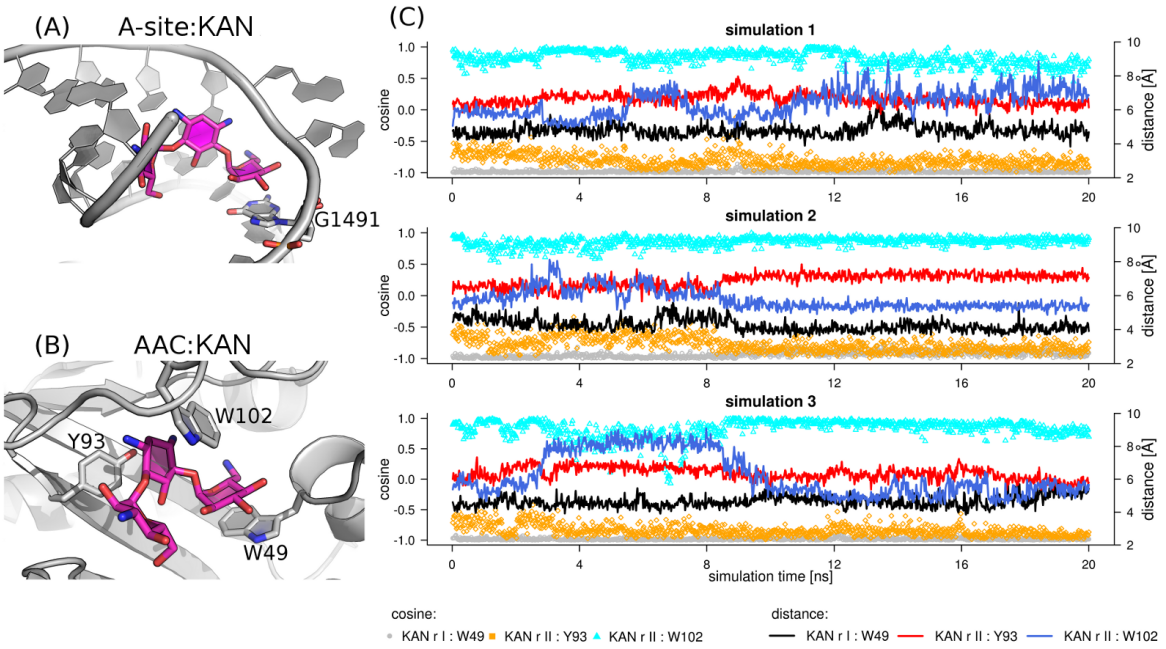


Figure 4.36: **Stacking interactions formed with kanamycin A in different receptors**. *(A)* Kanamycin A (magenta sticks) complexed with the A-site model; ring I of kanamycin is involved in stacking interactions with G1491 (shown as sticks). *(B)* A snapshot from the AAC_HOLO simulation showing the relative position of kanamycin A (magenta sticks) and the three amino acids involved in stacking (grey sticks). *(C)* The stacking interactions between AAC and kanamycin A, measured as: a cosine of the angle formed by two vectors perpendicular to the planes of the rings (plotted as points; a value close to 1 or -1 indicates a parallel orientation); and the distance between the rings involved in stacking (plotted as lines).

4.3.6 Water molecules inside aminoglycoside binding sites in AMEs

In the complexes of aminoglycosides with the RNA A-site, non-direct (water-mediated) interactions are often critical.[31,142] Although there were no such interactions reported in the crystal structure of RNA:KAN complex,[31] we monitored the water density in the binding clefts in the AME:KAN complexes. Additionally, since the simulations were performed with the all-atom representation of water molecules, we could investigate the mobility of these individual particles.

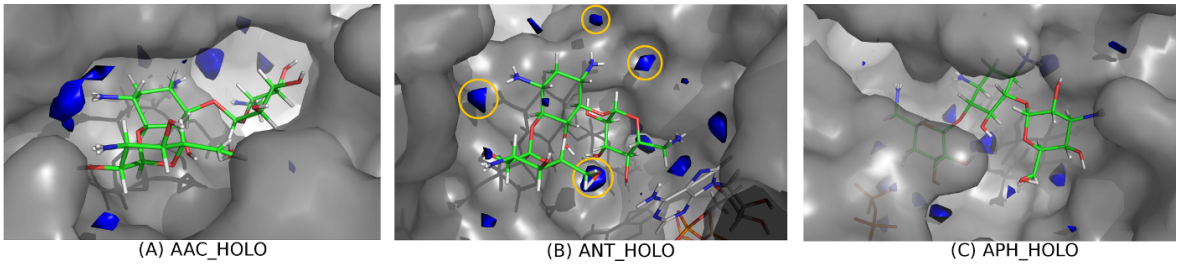


Figure 4.37: **Areas of high water density ($\geqslant$ 0.06, $\geqslant$ 0.07 and $\geqslant$ 0.085 water molecule/Å^3 for AAC_HOLO, ANT_HOLO, and APH_HOLO, respectively) around kanamycin A (in green) in the complexes with AMEs** shown as blue surfaces and superimposed on average structures calculated for one of the three independent simulations; the transparent surface represents the water accessible surface of the enzyme. The yellow circles mark the areas that were invariant in all ANT_HOLO simulations.

Water molecules in AME binding sites provide additional stability of the complexes

The average water density areas in the AME binding sites are presented in Figure 4.37. Different density values had to be set for each of the AME:KAN complexes — in AAC_HOLO even for values as low as 0.06 water molecules per Å^3 the areas were well defined, while in APH_HOLO and ANT_HOLO the visualization required higher water density thresholds. This demonstrates that in the AAC_HOLO simulation, the water molecules did not gather around the bound drug, in contrast to other AME simulations.

Inside the ANT binding site, many water dense areas were located between the atoms of the enzyme and the drug, which suggests that several indirect hydrogen bonds were formed in this complex. Moreover, since the antibiotic was quite dynamic inside the ANT cleft, the positions of these areas differed slightly in each binding site, with only a few conserved areas, which are marked in Figure 4.37B. In contrast, the areas of high water density in AAC_HOLO were farther away from kanamycin A, indicating a vacant space where another aminoglycoside of different shape could fit (Figure 4.37A).

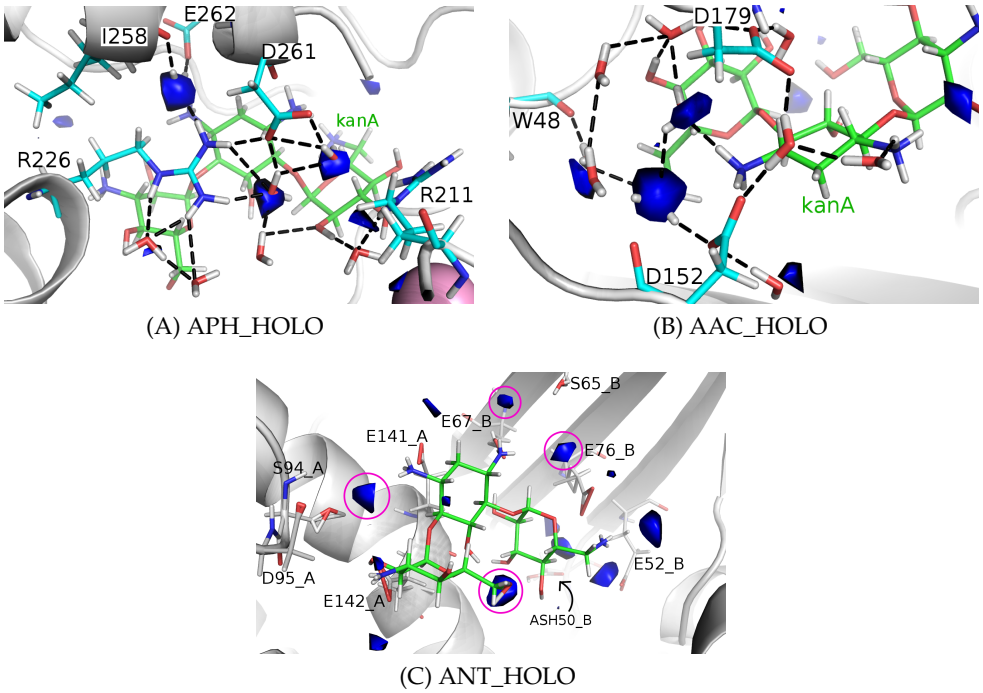
A few high water density areas in APH_HOLO were located between the drug and Arg 211, Arg 226, Ile 258, Asp 261, and Glu 262, indicating that a complicated network of hydrogen bonds was formed with these residues. Figure 4.38A presents an exemplary snapshot from the APH_HOLO simulation, which illustrates this network of hydrogen bonds. The water molecules that were part of these interactions interchanged with the bulk water, but the pattern of the hydrogen bonds was preserved throughout each of the three separate APH_HOLO simulations.

Similarly, in AAC_HOLO, some dense areas were positioned near ring II of the antibiotic and residues Asp 152, Asp 179 (Figure 4.38B). Also here, these dense areas symbolize a hydrogen bond network formed between these amino acids. As previously mentioned, in the third complex, ANT:KAN only a few water dense areas were invariant between different simulation runs and binding sites. These pointed to additional polar interactions formed by

the enzyme with kanamycin A, through the water molecules: *(i)* between the backbone moieties of Glu 141, Glu 142 and the oxygen O5 of KAN; *(ii)* between N1 of KAN and Glu 92, Ser 94, Asp 95; and *(iii)* between N3 of KAN and Glu 67, Glu 76 (Figure 4.38C).

In each simulation of the AME:KAN complexes, one or more water molecules were exceptionally stable, with the calculated β-factor $\leqslant$ 30 Å^2. While some of these "trapped" water molecules were positioned close to the bound drug, they never interacted with the drug directly, only assisted in positioning the nearby residues. In ANT_HOLO, only a few of the trapped water molecules were located near one of the kanamycin A molecules and they interacted with different amino acids in each simulation. Therefore, perhaps these water molecules were not significant for the complex formation. In the AAC_HOLO simulation, we found only one such stable water molecule, close to residues Tyr 93, Trp 103, Thr 151, Asp 152, Ile 114, and Asp 115 (Figure 4.39A). Interestingly, this was the same water molecule in each of the three independent runs of this simulation. In the last complex simulation, APH_HOLO, the trapped water molecules gathered close to the magnesium ions and helped to stabilize the nearby amino acids, Asp 208 and Asp 190 (Figure 4.39B).

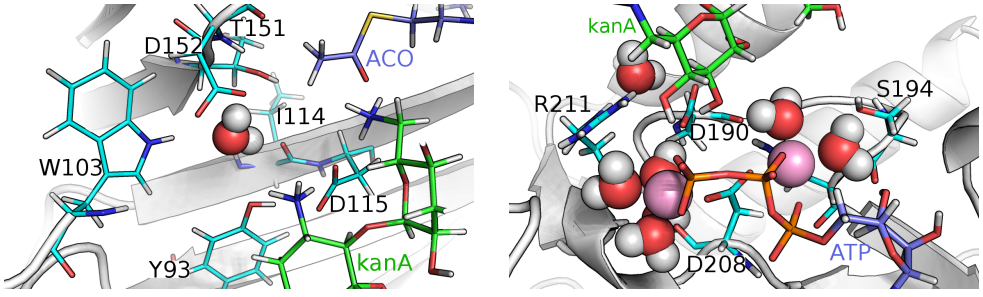
Although the strongest receptor:ligand interactions are those formed directly, water molecules are often found in binding sites of experimentally resolved complex structures. It has been shown that these water molecules buried deeply inside active sites of enzymes have different features from water in bulk.[224] Moreover, displacing a water molecule that bridges the receptor:ligand interaction and substituting it for a covalently bound moiety can lower the efficacy of binding.[225] Other theoretical studies highlight the importance of taking into account water-bridged interactions when analyzing structure activity relationships during a drug design process.[226] As described above, our simulations reveal potentially important sites for water-mediated interactions previously not explored.



(A) APH_HOLO (B) AAC_HOLO

(C) ANT_HOLO

Figure 4.38: **Water-mediated hydrogen bonds near kanamycin A in AME binding clefts.** The blue surfaces represent the areas of high water density ($\geqslant 0.06$, $\geqslant 0.07$ and $\geqslant 0.085$ water molecules per Å^3 for AAC_HOLO, ANT_HOLO, and APH_HOLO, respectively). In panels *(A–B)* the dense areas are superimposed on a snapshot from the simulations, which confirms that the shown water-bridged hydrogen bonds lasted for almost entire simulation. In panel *(C)* an average structure of ANT_HOLO is shown. The sticks are colored according to the atom type: H atoms are white, O — red, N — blue, C atoms of the antibiotic — green, other C atoms — cyan. The magnesium ion is shown as a pink sphere.



(A) AAC_HOLO (B) APH_HOLO

Figure 4.39: **The most stable water molecules in the AME:KAN complexes**. Snapshots from the simulations showing the interactions between the enzyme, kanamycin A and the "trapped" water molecules (in the case of APH_HOLO, β-factors for the water molecules for all three simulations were $\leqslant$ 30 Å^2; in the case of AAC_HOLO, mean β-factor for all three simulations was 13.5 ± 2.3 Å^2). The water molecules are shown as spheres, and the interacting residues — as sticks colored according to atom type: H atoms in white, O — red, N — blue, P — orange, C atoms of the antibiotic — green, C atoms of ATP or ACO — light blue, other C atoms — cyan. The magnesium ions are shown as pink spheres.

4.3.7 Energetical analysis of kanamcyin A complexes

Enzymes mimic predominantly electrostatic interactions

A per-residue decomposition of the MMGBSA binding free energies was performed to show the most favorable interactions formed with kanamycin A in different receptors. In Figure 4.40 the five amino acids contributing the most to binding with kanamycin A are illustrated and their energetic contribution is decomposed into parts that come from the van der Waals (vdW), electrostatic (e-s) or non-polar term (related to the solvent accessible surface, sas). In the rRNA binding site, the phosphate backbone created the most favorable interactions (phosphate groups of G1494 and A1493). However, the enzymes utilized the *side chains* of amino acids to mimic these interactions. As mentioned earlier, different amino acids were involved in hydrogen bonding with kanamycin A, but the energetic analysis shows that the most important were Asp or Glu. In the A-site every aminoglycoside has its ring I stacked on top of G1491, which suggests that it is an important interaction. However, the energetical analysis shows that this base did not give a significant contribution to binding (the van der Waals component of the energy was relatively strong, but the electrostatic part did not favor binding; data not shown). This can possibly provide an explanation for the majority of AMEs not reproducing the stacking inside their binding cleft. It could be more efficient to focus on mimicking the electrostatic interactions. Moreover, amino acids forming stacking have large hydrophobic side chains, which cannot change their conformation as easily as e.g., Asp or Glu, and this flexibility is necessary if an enzyme is to adapt to different shapes and sizes of aminoglycosides.

The energetical analysis showed that the Glu residues from the aminoglycoside binding loop in APH (residues from around Asp 150 to Asp 162) were the most important for kanamycin A binding (Figure 4.40). In contrast, the amino acids forming the mobile loop in AAC (from Glu 47 to Arg 55) did not give such a strong contribution to the binding free energy — in the five most significant amino acids, only Trp 49 is present, contributing strongly through the stacking. This suggests that the role of this AAC fragment may be to block a premature release of the drug from the binding pocket, while in APH, the loop specifically recognizes the antibiotic. Importantly, the relative orientation of kanamycin A and the surface of the binding pocket in APH is rotated by 180° when compared with the other two enzymes or the A-site, as can be seen in the bottom panels in Figure 4.40. This means that the most energetically favorable interactions were formed predominantly with rings II and III in the case of the APH_HOLO simulations, and with rings I and II in the case of the other receptors. Nevertheless, rings I and II were forming strong hydrogen bonds with APH, as shown earlier. Another uniqe feature of APH_HOLO is the high electrostatic contribution to binding of the C-terminal residues Asp 261, Glu 262 and Phe 264. These residues were forming hydrogen bonds with kanamycin A using both, their side chains and backbone, as shown by the decomposition of the energy in Figure 4.40 and by the hydrogen bond analysis in Table 4.9. These interactions are most likely the same in complexes with different aminoglycosides, because these amino acids are highly conserved within the APH(3') subfamily and important for the activity of these enzymes.[222] The energetical analysis of the ANT:KAN complex is in agreement with recent mutational studies of ANT(4'),[220] which showed that the ANT double mutant E145Q/E76Q lacks enzyme activity, while the single mutants (E145Q and E76Q) had no substantial change in the reaction constants for modification of kanamycin A. Our energy decomposition demonstrated that both of these amino acids contribute favorably to the binding, and thus the lack of only one of these contributions would not have a big impact on the complex stability.

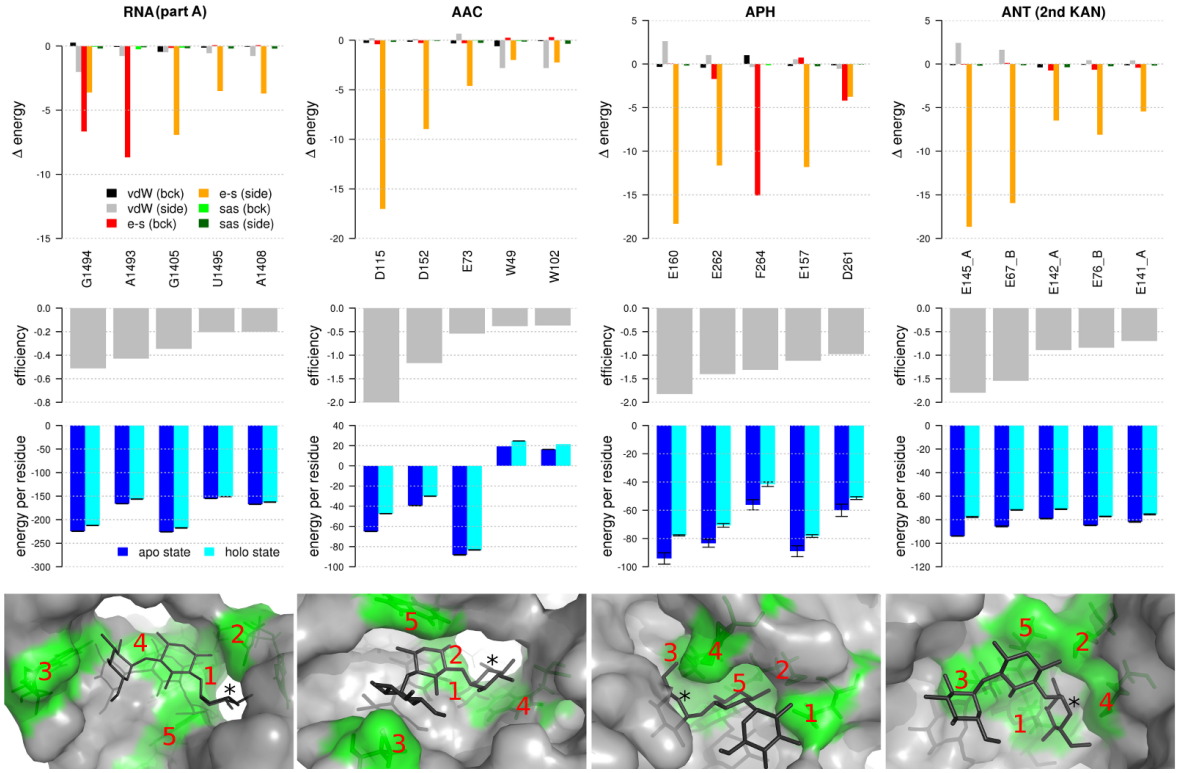


Figure 4.40: **Per-residue decomposition of the MMGBSA free binding energy for the studied AME complexes** (all energy values are given in kcal/mol).
(three top panels) "Efficiency" was calculated as the ratio of the MMGBSA energy difference to the number of heavy atoms in the residue and only the five most important energetically residues are shown in order of decreasing significance. The MMGBSA energy difference (Δ energy) between the calculated energy of a residue in the APO and HOLO simulations, is decomposed into: vdW = van der Waals interactions, e-s = electrostatic interactions, sas = non-polar contribution, related to the solvent accessible surface area; abbreviations in brackets: bck = backbone, side = side chain.
(bottom panels) The residues that favor kanamycin A binding (black sticks) are marked green and labeled in the order of importance; the asterisks (*) mark ring I — notice that the orientation of kanamycin A in the APH binding cleft is opposite to the one in the A-site.

4.3.8 Summary

The results from our simulations of AME representatives highlighted the similarities and differences between the aminoglycoside binding sites in these enzymes and in the rRNA A-site. The three enzymes were found to have different flexibility patterns, which we hypothesize correlates with different ways of achieving promiscuity towards aminoglycosides. The APH representative was the most flexible in the APO state, and it only stabilized upon complex formation with the drug. The different conformations sampled by this enzyme show that the binding of the drug occurs through an induced fit mechanism. The AAC enzyme was also flexible, mostly due to high mobility of the loops located over the aminoglycoside binding site. This finding is similar to what we observed in the APH simulations; however, in contrast to APH, the loop in AAC did not form energetically important interactions with the drug. We postulate that AAC is able to bind different aminoglycosides because it possesses aromatic residues in the binding site that form stacking interactions with the aminoglycoside rings. The third enzyme, ANT, was the least flexible, but its binding site is comprised of many hydrogen bond donors that create direct and water-mediated interactions with aminoglycosides.

Our analysis demonstrated how similar the interactions formed by the enzymes and RNA with kanamycin A were. Although each enzyme has a slightly different electrostatic potential inside its binding site, the hydrogen bonds and even stacking were literally copied from the A-site:KAN complex. Energetic analysis confirmed that the most important interactions formed with kanamycin A were of the same nature, with Asp or Glu residues mimicking the phosphate groups of RNA. In addition, we showed that water molecules are involved in stabilizing AME:KAN complexes, through a complicated network of hydrogen bonds with the amino acids inside the aminoglycoside binding sites.

Chapter 5

Conclusions and outlook

The research presented here focused on clarifying the atomic details of two mechanisms of bacterial resistance against aminoglycoside antibiotics: *(i)* mutations in the region of the primary aminoglycoside binding site, the A-site in 16S rRNA; and *(ii)* chemical modification of aminoglycosides by AMEs. The computational biophysical tools enabled us to study the static and dynamic features of the biomolecules that bind aminoglycosides inside bacterial cell. Additionally, we proposed a new tool that analyzes the conformations of a molecule and defines its division into dynamic domains. This approach facilitates the description of dynamic changes that can be sampled through NMR experiments or computer simulations.

In the first part of the research different mutations in the A-site were examined. We showed that these substitutions have different effects on the physicochemical features of this RNA fragment. Figure 5.1 illustrates this a concise manner. We have found that some mutations affected the shape of the aminoglycoside binding site, while some changed the charge distribution. Our results are in accord with experimental findings and suggested which aminoglycoside substituents should be modified so that aminoglycosides would also bind to the mutated A-site to overcome mutation-related bacterial resistance.

Concerning the second bacterial resistance method, we have found both, similar and dissimilar features of the aminoglycoside binding sites in representative AMEs and RNA. We noted that the three AME representatives employ different techniques to effectively bind aminoglycosides that vary in shape and size. On the other hand, it appears that AMEs, built of amino acids, mimic the interaction sites from RNA, which is made of nucleic bases (illustrated in Figure 5.2). The energetic analysis showed that these interactions were copied efficiently.

The main tool used here, MD simulation, enables detailed analysis of atomic interactions; however, due to its approximations, it has many drawbacks. The feature that most influences the results is the limited sampling of one MD simulation, which is restricted to one or a few local energy minima. Therefore, we tried to increase the sampling by either simulating two copies of a system or performing multiple simulation runs. However, even this cannot guarantee observation of all conformations that are important for a given problem. Next, the force field is only a molecular-mechanics based approximation of a real description of interatomic interactions, hence the potential energies of a molecule and the free binding energies give just a crude estimation of the real values and only their relative values can be taken into account. In addition, the time scales of conformational changes do not mirror the experimental values, and therefore, the observations have only statistical meaning. Moreover, even though the force fields have significantly improved since their first usage in the 1970s, they are still far from being perfect. This is especially the case for RNA because the RNA-specific force fields were formulated only some time after the force fields for proteins. Nevertheless, when used with caution, MD simulation is a very powerful tool.

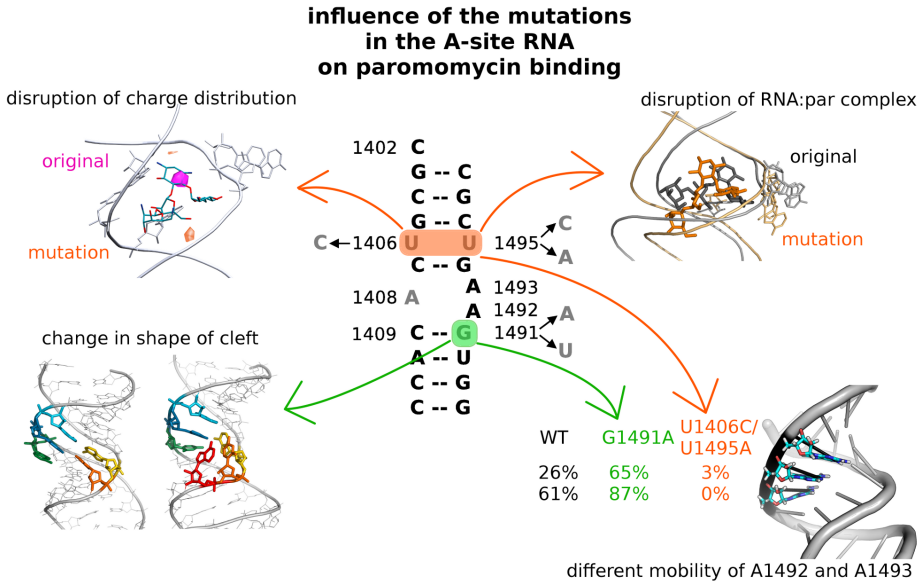


Figure 5.1: **Graphical representation of key findings concerning mutations in the rRNA A-site.** The mutations studied here led to different changes in the physicochemical features of the binding site: the biggest changes were caused by the double mutation U1406C/U1495A, including disruption of the hydrogen bonds formed with its complex with paromomycin; mutations in position 1491 altered the shape of the binding cleft; all mutations affected the mobility of A1492 and A1493 (shown here are exemplary values presenting the percentage of the simulation time when each adenine was in the flipped-in state).

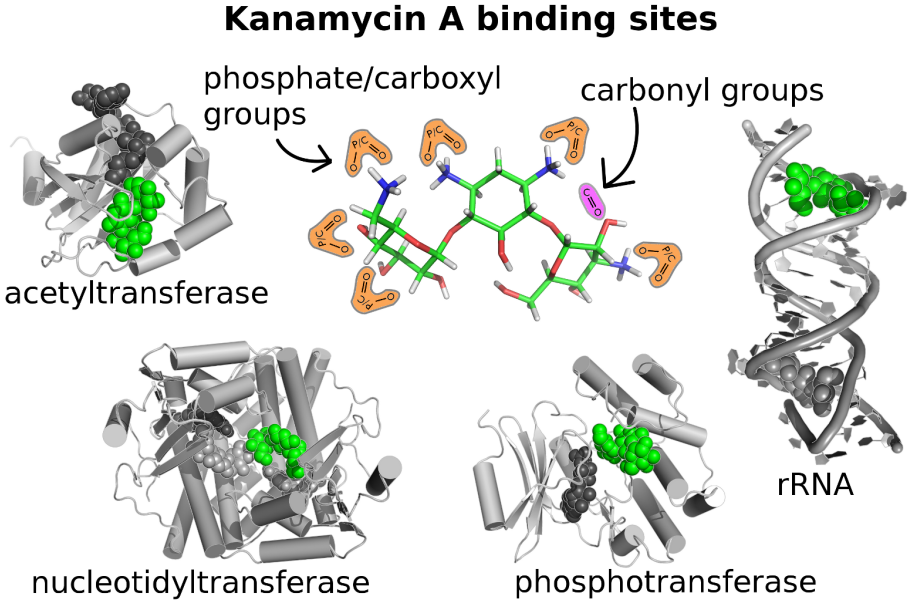


Figure 5.2: **Graphical representation of key findings concerning AME-mediated aminoglycoside modifications.** Kanamycin A is recognized and bound by different biomolecules: *(from top right, anti-clockwise)* the rRNA A-site, the very flexible APH(3′,5″)-IIIa, the dimeric and relatively rigid ANT(4′), and the flexible AAC(6′)-Ib. Despite the intrinsic differences in the entities that compose these receptors (nucleotides in RNA and amino acids in AMEs), these various biomolecules create very similar environment inside the aminoglycoside binding sites, utilizing mainly electrostatic interactions to effectively bind aminoglycosides.

The results presented here may lead to some aminoglycoside modifications that would enable them to bind to the rRNA A-site but hinder binding to AMEs. Then, a proper structure-based drug design approach could be employed to yield several compounds that could be tested in the laboratory. Apart from application to antibiotic design, the AME superfamily presents an elegant system for studying the relation between structure and function of a protein. The enzymes from the different families, AACs, ANTs, and APHs, differ in structure, but still perform the same function. Moreover, the AAC(6′)-Ib studied here has a unique sequence but similar structure to other AAC(6′) enzymes. The modeling methods and comparative analysis of more AME:drug complexes could give insight into specific features that must be present in order for different structures to perform a similar function. In addition, aminoglycosides and their receptors, especially RNA, are a strong test system for improving existing docking procedures. Not only must the drugs be adjusted to the receptor, but both the RNA and AMEs have different conformations when bound to different ligands, which was also shown here. Therefore, development of the flexible docking procedures is needed. Moreover, none of the well-established docking programs account for water molecules in a binding site, which can significantly increase the stability of a complex.

Finally, we presented a new algorithm for analyzing conformational changes in biomolecules, GeoStaS. We have shown that our new approach performs better than a commonly used tool, DCCM, to analyze MD simulations. Our algorithm and software correctly recognizes correlated motions and can facilitate the analysis of large trajectories through visualization of the identified dynamic domains. However, the implementation still needs some improvement. We plan to test another method of clustering AMSM, and improve the process of choosing the optimal division into dynamic domains. Additionally, to complete the description of the conformational diversity of a molecule, we would like to find a representative set of conformations, i.e., cluster a trajectory. These transitions can then be presented as relations between the clusters, and the relative orientations of the dynamic domains can efficiently describe the clusters. Such an approach would enable researchers to focus on important details of the molecular interactions rather than on the time-consuming post-processing of a trajectory.

Bibliography

[1] Sköld, O. (2011) *Antibiotics and Antibiotic Resistance*. Wiley.

[2] Eagon, R. G. (1962) Pseudomonas natriegens, a marine bacterium with a generation time of less than 10 minutes. *J Bacteriol*, **83**, 736–7.

[3] D'Costa, V. M., et al. (2011) Antibiotic resistance is ancient. *Nature*, **477**, 457–461.

[4] Ditzler, M. A., Otyepka, M., Šponer, J., and Walter, N. G. (2010) Molecular Dynamics and Quantum Mechanics of RNA: Conformational and Chemical Change We Can Believe In. *Accounts Chem Res*, **43**, 40–47.

[5] Šponer, J. and Lankaš, F. (eds.) (2006) *Computational studies of RNA and DNA*. Springer.

[6] Schneider, G. and Baringhaus, K.-H. (2008) *Molecular Design: Concepts and Applications*. Wiley-VCH.

[7] Caminero, J. A., Sotgiu, G., Zumla, A., and Migliori, G. B. (2010) Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis. *Lancet Infect Dis*, **10**, 621–9.

[8] World Health Organization (2010) Multidrug and extensively drug-resistant TB (M/XDR-TB). Global Report on Surveillance and Response. Tech. rep., World Health Organization.

[9] Walsh, C. (2003) *Antibiotics: Actions, Origins, Resistance*. Amer Society for Microbiology.

[10] Vakulenko, S. B. and Mobashery, S. (2003) Versatility of aminoglycosides and prospects for their future. *Clin Microbiol Rev*, **16**, 430–50.

[11] Jana, S. and Deb, J. K. (2006) Molecular understanding of aminoglycoside action and resistance. *Appl Microbiol Biot*, **70**, 140–50.

[12] Kaul, M., Barbieri, C. M., Kerrigan, J. E., and Pilch, D. S. (2003) Coupling of Drug Protonation to the Specific Binding of Aminoglycosides to the A Site of 16S rRNA: Elucidation of the Number of Drug Amino Groups Involved and their Identities. *J Mol Biol*, **326**, 1373–1387.

[13] Ramakrishnan, V. (2002) Ribosome structure and the mechanism of translation. *Cell*, **108**, 557–72.

[14] Magnet, S. and Blanchard, J. S. (2005) Molecular insights into aminoglycoside action and resistance. *Chem Rev*, **105**, 477–98.

[15] Tselika, S., Konstantinidis, T. C., and Synetos, D. (2008) Two nucleotide substitutions in the A-site of yeast 18S rRNA affect translation and differentiate the interaction of ribosomes with aminoglycoside antibiotics. *Biochimie*, **90**, 908–17.

[16] Pfister, P., Risch, M., Brodersen, D. E., and Böttger, E. C. (2003) Role of 16S rRNA Helix 44 in Ribosomal Resistance to Hygromycin B. *Antimicrob Agents Ch*, **47**, 1496–502.

[17] Hobbie, S. N., Akshay, S., Kalapala, S. K., Bruell, C. M., Shcherbakov, D., and Böttger, E. C. (2008) Genetic analysis of interactions with eukaryotic rRNA identify the mitoribosome as target in aminoglycoside ototoxicity. *P Natl Acad Sci USA*, **105**, 20888–93.

[18] Fourmy, D., Yoshizawa, S., and Puglisi, J. D. (1998) Paromomycin binding induces a local conformational change in the A-site of 16 S rRNA. *J Mol Biol*, **277**, 333–45.

[19] Lynch, S. R. and Puglisi, J. D. (2001) Structure of a eukaryotic decoding region A-site RNA. *J Mol Biol*, **306**, 1023–35.

[20] Meroueh, S. O. and Mobashery, S. (2007) Conformational transition in the aminoacyl t-RNA site of the bacterial ribosome both in the presence and absence of an aminoglycoside antibiotic. *Chem Biol Drug Des*, **69**, 291–7.

[21] Fan-Minogue, H. and Bedwell, D. M. (2008) Eukaryotic ribosomal RNA determinants of aminoglycoside resistance and their role in translational fidelity. *RNA*, **14**, 148–57.

[22] Laurberg, M., Asahara, H., Korostelev, A., Zhu, J., Trakhanov, S., and Noller, H. F. (2008) Structural basis for translation termination on the 70S ribosome. *Nature*, **454**, 852–7.

[23] Korostelev, A., Asahara, H., Lancaster, L., Laurberg, M., Hirschi, A., Zhu, J., Trakhanov, S., Scott, W. G., and Noller, H. F. (2008) Crystal structure of a translation termination complex formed with release factor RF2. *P Natl Acad Sci USA*, **105**, 19684–9.

[24] McClory, S. P., Leisring, J. M., Qin, D., and Fredrick, K. (2010) Missense suppressor mutations in 16S rRNA reveal the importance of helices h8 and h14 in aminoacyl-tRNA selection. *RNA*, **16**, 1925–34.

[25] Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*, **407**, 340–8.

[26] Shandrick, S., Zhao, Q., Han, Q., Ayida, B. K., Takahashi, M., Winters, G. C., Simonsen, K. B., Vourloumis, D., and Hermann, T. (2004) Monitoring molecular recognition of the ribosomal decoding site. *Angew Chem Int Edit*, **43**, 3177–82.

[27] Chao, P.-W. and Chow, C. S. (2007) Monitoring aminoglycoside-induced conformational changes in 16S rRNA through acrylamide quenching. *Bioorg Med Chem*, **15**, 3825–31.

[28] Houghton, J. L., Green, K. D., Chen, W., and Garneau-Tsodikova, S. (2010) The Future of Aminoglycosides: The End or Renaissance? *ChemBioChem*, **11**, 880–902.

[29] Nudelman, I., Rebibo-Sabbah, A., Shallom-Shezifi, D., Hainrichson, M., Stahl, I., Ben-Yosef, T., and Baasov, T. (2006) Redesign of aminoglycosides for treatment of human genetic diseases caused by premature stop mutations. *Bioorg Med Chem Lett*, **16**, 6310–5.

[30] Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature*, **407**, 327–39.

[31] François, B., Russell, R. J. M., Murray, J. B., Aboul-ela, F., Masquida, B., Vicens, Q., and Westhof, E. (2005) Crystal structures of complexes between aminoglycosides and decoding A site oligonucleotides: role of the number of rings and positive charges in the specific binding leading to miscoding. *Nucl Acids Res*, **33**, 5677–90.

[32] Xi, H., Kumar, S., Dosen-Micovic, L., and Arya, D. P. (2010) Calorimetric and spectroscopic studies of aminoglycoside binding to AT-rich DNA triple helices. *Biochimie*, **92**, 514–529.

[33] Schroeder, R., Waldsich, C., and Wank, H. (2000) Modulation of RNA function by aminoglycoside antibiotics. *EMBO J*, **19**, 1–9.

[34] Recht, M. I., Fourmy, D., Blanchard, S. C., Dahlquist, K. D., and Puglisi, J. D. (1996) RNA sequence determinants for aminoglycoside binding to an A-site rRNA model oligonucleotide. *J Mol Biol*, **262**, 421–36.

[35] Anderson, P. C. and Mecozzi, S. (2007) Minimum sequence requirements for the binding of paromomycin to the rRNA decoding site A. *Biopolymers*, **86**, 95–111.

[36] Tran, T. and Disney, M. D. (2010) Two-dimensional combinatorial screening of a bacterial rRNA A-site-like motif library: defining privileged asymmetric internal loops that bind aminoglycosides. *Biochemistry*, **49**, 1833–42.

[37] Hermann, T. and Westhof, E. (1998) RNA as a drug target: chemical, modelling, and evolutionary tools. *Curr Opin Biotech*, **9**, 66–73.

[38] Tor, Y. (2006) The ribosomal A-site as an inspiration for the design of RNA binders. *Biochimie*, **88**, 1045–51.

[39] Tor, Y. and Westhof, E. (1998) Deciphering RNA recognition: the hammerhead ribozyme. *Chem Biol*, **5**, R277–R283.

[40] Lorian, V. (ed.) (2005) *Antibiotics in Laboratory Medicine*. Lippincott Williams & Wilkins.

[41] Hobbie, S. N., Pfister, P., Brüll, C., Westhof, E., and Böttger, E. C. (2005) Analysis of the contribution of individual substituents in 4,6-aminoglycoside-ribosome interaction. *Antimicrob Agents Ch*, **49**, 5112–8.

[42] Hobbie, S. N., Pfister, P., Bruell, C., Sander, P., François, B., Westhof, E., and Böttger, E. C. (2006) Binding of neomycin-class aminoglycoside antibiotics to mutant ribosomes with alterations in the A site of 16S rRNA. *Antimicrob Agents Ch*, **50**, 1489–96.

[43] De Stasio, E. A. and Dahlberg, A. E. (1990) Effects of mutagenesis of a conserved base-paired site near the decoding region of Escherichia coli 16 S ribosomal RNA. *J Mol Biol*, **212**, 127–33.

[44] Recht, M. I., Douthwaite, S., Dahlquist, K. D., and Puglisi, J. D. (1999) Effect of mutations in the A site of 16 S rRNA on aminoglycoside antibiotic-ribosome interaction. *J Mol Biol*, **286**, 33–43.

[45] Hobbie, S. N., Bruell, C., Kalapala, S., Akshay, S., Schmidt, S., Pfister, P., and Böttger, E. C. (2006) A genetic model to investigate drug-target interactions at the ribosomal decoding site. *Biochimie*, **88**, 1033–43.

[46] Gregory, S. T., Carr, J. F., Rodriguez-Correa, D., and Dahlberg, A. E. (2005) Mutational analysis of 16S and 23S rRNA genes of Thermus thermophilus. *J Bacteriol*, **187**, 4804–12.

[47] Vila-Sanjurjo, A., Lu, Y., Aragonez, J. L., Starkweather, R. E., Sasikumar, M., and O'Connor, M. (2007) Modulation of 16S rRNA function by ribosomal protein S12. *Biochim Biophys Acta*, **1769**, 462–71.

[48] Pfister, P., Hobbie, S. N., Brüll, C., Corti, N., Vasella, A., Westhof, E., and Böttger, E. C. (2005) Mutagenesis of 16S rRNA C1409-G1491 base-pair differentiates between 6'OH and 6'NH3+ aminoglycosides. *J Mol Biol*, **346**, 467–75.

[49] Pfister, P., Hobbie, S. N., Vicens, Q., Böttger, E. C., and Westhof, E. (2003) The molecular basis for A-site mutations conferring aminoglycoside resistance: relationship between ribosomal susceptibility and X-ray crystal structures. *Chembiochem*, **4**, 1078–88.

[50] Lynch, S. R. and Puglisi, J. D. (2001) Structural origins of aminoglycoside specificity for prokaryotic ribosomes. *J Mol Biol*, **306**, 1037–58.

[51] Shcherbakov, D., Akbergenov, R., Matt, T., Sander, P., Andersson, D. I., and Böttger, E. C. (2010) Directed mutagenesis of Mycobacterium smegmatis 16S rRNA to reconstruct the in-vivo evolution of aminoglycoside resistance in Mycobacterium tuberculosis. *Mol Microbiol*, **77**, 830–840.

[52] Corzana, F., Cuesta, I., Freire, F., Revuelta, J., Torrado, M., Bastida, A., Jiménez-Barbero, J., and Asensio, J. L. (2007) The pattern of distribution of amino groups modulates the structure and dynamics of natural aminoglycosides: implications for RNA recognition. *J Am Chem Soc*, **129**, 2849–2865.

[53] Doi, Y. and Arakawa, Y. (2007) 16S ribosomal RNA methylation: emerging resistance mechanism against aminoglycosides. *Clin Infect Dis*, **45**, 88–94.

[54] Kimura, S. and Suzuki, T. (2010) Fine-tuning of the ribosomal decoding center by conserved methyl-modifications in the Escherichia coli 16S rRNA. *Nucl Acids Res*, **38**, 1341–1352.

[55] Ramirez, M. S. and Tolmasky, M. E. (2010) Aminoglycoside modifying enzymes. *Drug Resist Update*, **13**, 151–71.

[56] Llano-Sotelo, B., Azucena, E. F., Kotra, L. P., Mobashery, S., and Chow, C. S. (2002) Aminoglycosides modified by resistance enzymes display diminished binding to the bacterial ribosomal aminoacyl-tRNA site. *Chem Biol*, **9**, 455–463.

[57] Vetting, M. W., de Carvalho, L. P. S., Yu, M., Hegde, S. S., Magnet, S., Roderick, S. L., and Blanchard, J. S. (2005) Structure and functions of the GNAT superfamily of acetyltransferases. *Arch Biochem Biophys*, **433**, 212–26.

[58] Daigle, D. M., McKay, G. A., Thompson, P. R., and Wright, G. D. (1999) Aminoglycoside antibiotic phosphotransferases are also serine protein kinases. *Chem Biol*, **6**, 11–8.

[59] Lombès, T., Bégis, G., Maurice, F., Turcaud, S., Lecourt, T., Dardel, F., and Micouin, L. (2008) NMR-guided fragment-based approach for the design of AAC(6′)-Ib ligands. *ChemBioChem*, **9**, 1368–71.

[60] Vetting, M. W., Park, C. H., Hegde, S. S., Jacoby, G. A., Hooper, D. C., and Blanchard, J. S. (2008) Mechanistic and structural analysis of aminoglycoside N-acetyltransferase AAC(6′)-Ib and its bifunctional, fluoroquinolone-active AAC(6′)-Ib-cr variant. *Biochemistry*, **47**, 9825–35.

[61] Rather, P. N., Munayyer, H., Mann, P. A., Hare, R. S., Miller, G. H., and Shaw, K. J. (1992) Genetic analysis of bacterial acetyltransferases: identification of amino acids determining the specificities of the aminoglycoside 6′-N-acetyltransferase Ib and IIa proteins. *J Bacteriol*, **174**, 3196–203.

[62] Shmara, A., Weinsetel, N., Dery, K. J., Chavideh, R., and Tolmasky, M. E. (2001) Systematic analysis of a conserved region of the aminoglycoside 6′-N-acetyltransferase type Ib. *Antimicrob Agents Ch*, **45**, 3287–92.

[63] Casin, I., Hanau-Berçot, B., Podglajen, I., Vahaboglu, H., and Collatz, E. (2003) Salmonella enterica serovar Typhimurium bla(PER-1)-carrying plasmid pSTI1 encodes an extended-spectrum aminoglycoside 6′-N-acetyltransferase of type Ib. *Antimicrob Agents Ch*, **47**, 697–703.

[64] Pourreza, A., Witherspoon, M., Fox, J., Newmark, J., Bui, D., and Tolmasky, M. E. (2005) Mutagenesis analysis of a conserved region involved in acetyl coenzyme A binding in the aminoglycoside 6′-N-acetyltransferase type Ib encoded by plasmid pJHCMW1. *Antimicrob Agents Ch*, **49**, 2979–82.

[65] Burk, D. L., Ghuman, N., Wybenga-Groot, L. E., and Berghuis, A. M. (2003) X-ray structure of the AAC(6′)-Ii antibiotic resistance enzyme at 1.8 A resolution; examination of oligomeric arrangements in GNAT superfamily members. *Protein Sci*, **12**, 426–437.

[66] Sakon, J., Liao, H. H., Kanikula, A. M., Benning, M. M., Rayment, I., and Holden, H. M. (1993) Molecular structure of kanamycin nucleotidyltransferase determined to 3 Å resolution. *Biochemistry*, **32**, 11977–11984.

[67] Revuelta, J., Vacas, T., Torrado, M., Corzana, F., Gonzalez, C., Jiménez-Barbero, J., Menendez, M., Bastida, A., and Asensio, J. L. (2008) NMR-based analysis of aminoglycoside recognition by the resistance enzyme ANT(4′): the pattern of OH/NH3(+) substitution determines the preferred antibiotic binding mode and is critical for drug inactivation. *J Am Chem Soc*, **130**, 5086–103.

[68] Pedersen, L. C., Benning, M. M., and Holden, H. M. (1995) Structural investigation of the antibiotic and ATP-binding sites in kanamycin nucleotidyltransferase. *Biochemistry*, **34**, 13305–13311.

[69] Thompson, P. R., Hughes, D. W., and Wright, G. D. (1996) Mechanism of aminoglycoside 3′-phosphotransferase type IIIa: His188 is not a phosphate-accepting residue. *Chem Biol*, **3**, 747–55.

[70] Boehr, D. D., Thompson, P. R., and Wright, G. D. (2001) Molecular mechanism of aminoglycoside antibiotic kinase APH(3′)-IIIa: roles of conserved active site residues. *J Biol Chem*, **276**, 23929–36.

[71] Thompson, P. R., Boehr, D. D., Berghuis, A. M., and Wright, G. D. (2002) Mechanism of aminoglycoside antibiotic kinase APH(3′)-IIIa: role of the nucleotide positioning loop. *Biochemistry*, **41**, 7001–7.

[72] Wu, L. and Serpersu, E. H. (2009) Deciphering Interactions of the Aminoglycoside Phosphotransferase(3′)- IIIa with its Ligands. *Biopolymers*, **91**, 801–809.

[73] Burk, D. L., Hon, W. C., Leung, A. K., and Berghuis, A. M. (2001) Structural analyses of nucleotide binding to an aminoglycoside phosphotransferase. *Biochemistry*, **40**, 8756–64.

[74] Fong, D. H. and Berghuis, A. M. (2002) Substrate promiscuity of an aminoglycoside antibiotic resistance enzyme via target mimicry. *EMBO J*, **21**, 2323–2331.

[75] Norris, A. L. and Serpersu, E. H. (2009) NMR detected hydrogen-deuterium exchange reveals differential dynamics of antibiotic- and nucleotide-bound aminoglycoside phosphotransferase 3′-IIIa. *J Am Chem Soc*, **131**, 8587–94.

[76] Cox, J. R., McKay, G. A., Wright, G. D., and Serpersu, E. H. (1996) Arrangement of Substrates at the Active Site of an Aminoglycoside Antibiotic 3′-Phosphotransferase As Determined by NMR. *J Am Chem Soc*, **118**, 1295–1301.

[77] Cox, J. R. and Serpersu, E. H. (1997) Biologically important conformations of aminoglycoside antibiotics bound to an aminoglycoside 3′-phosphotransferase as determined by transferred nuclear Overhauser effect spectroscopy. *Biochemistry*, **36**, 2353–2359.

[78] Cox, J. R., Ekman, D. R., DiGiammarino, E. L., Akal-Strader, A., and Serpersu, E. H. (2000) Aminoglycoside antibiotics bound to aminoglycoside-detoxifying enzymes and RNA adopt similar conformations. *Cell Biochem Biophys*, **33**, 297–308.

[79] Serpersu, E. H., Cox, J. R., DiGiammarino, E. L., Mohler, M. L., Ekman, D. R., Akal-Strader, A., and Owston, M. (2000) Conformations of antibiotics in active sites of aminoglycoside-detoxifying enzymes. *Cell Biochem Biophys*, **33**, 309–21.

[80] Bastida, A., et al. (2006) Exploring the use of conformationally locked aminoglycosides as a new strategy to overcome bacterial resistance. *J Am Chem Soc*, **128**, 100–16.

[81] Hermann, T. and Westhof, E. (1999) Docking of cationic antibiotics to negatively charged pockets in RNA folds. *J Med Chem*, **42**, 1250–61.

[82] Réblová, K., Lankaš, F., Rázga, F., Krasovska, M. V., Koča, J., and Šponer, J. (2006) Structure, dynamics, and elasticity of free 16S rRNA helix 44 studied by molecular dynamics simulations. *Biopolymers*, **82**, 504–20.

[83] Sanbonmatsu, K. Y. (2006) Energy landscape of the ribosomal decoding center. *Biochimie*, **88**, 1053–9.

[84] Vaiana, A. C., Westhof, E., and Auffinger, P. (2006) A molecular dynamics simulation study of an aminoglycoside/A-site RNA complex: conformational and hydration patterns. *Biochimie*, **88**, 1061–73.

[85] Romanowska, J., Setny, P., and Trylska, J. (2008) Molecular dynamics study of the ribosomal A-site. *J Phys Chem B*, **112**, 15227–43.

[86] Chen, S.-Y. and Lin, T.-H. (2010) A molecular dynamics study on binding recognition between several 4,5 and 4,6-linked aminoglycosides with A-site RNA. *J Mol Recognit*, **23**, 423–34.

[87] Vaiana, A. C. and Sanbonmatsu, K. Y. (2009) Stochastic gating and drug-ribosome interactions. *J Mol Biol*, **386**, 648–61.

[88] Długosz, M., Antosiewicz, J. M., and Trylska, J. (2008) Association of aminoglycosidic antibiotics with the ribosomal A-site studied with Brownian dynamics. *J Chem Theory Comput*, **4**, 549–559.

[89] Długosz, M. and Trylska, J. (2009) Aminoglycoside association pathways with the 30S ribosomal subunit. *J Phys Chem B*, **113**, 7322–30.

[90] Yang, G., Trylska, J., Tor, Y., and McCammon, J. A. (2006) Binding of aminoglycosidic antibiotics to the oligonucleotide A-site model and 30S ribosomal subunit: Poisson-Boltzmann model, thermal denaturation, and fluorescence studies. *J Med Chem*, **49**, 5478–90.

[91] Krishna, V., Kumar, K. S. V., Anishetty, S., and Pennathur, G. (2009) Investigations on Domain Movements of N-Acetyltransferase in Nano Scale. *J Nanosci Nanotechno*, **9**, 5493–5496.

[92] Hu, X., Norris, A. L., Baudry, J. Y., and Serpersu, E. H. (2011) Coenzyme-A Binding to the Aminoglycoside Acetyltransferase (3)-IIIb Increases Conformational Sampling of Antibiotic Binding Site. *Biochemistry*, **50**, 10559–10565.

[93] Infectious Diseases Society of America (2010) The 10 x '20 Initiative: pursuing a global commitment to develop 10 new antibacterial drugs by 2020. *Clin Infect Dis*, **50**, 1081–3.

[94] Herper, M. (2012) The Truly Staggering Cost Of Inventing New Drugs. *Forbes*.

[95] Zhou, J., Wang, G., Zhang, L.-H., and Ye, X.-S. (2007) Modifications of aminoglycoside antibiotics targeting RNA. *Med Res Rev*, **27**, 279–316.

[96] Hotta, K., Sunada, A., Ikeda, Y., and Kondo, S. (2000) Double stage activity in aminoglycoside antibiotics. *J Antibiot*, **53**, 1168–74.

[97] Shaul, P., Green, K. D., Rutenberg, R., Kramer, M., Berkov-Zrihen, Y., Breiner-Goldstein, E., Garneau-Tsodikova, S., and Fridman, M. (2011) Assessment of 6'- and 6"'-N-acylation of aminoglycosides as a strategy to overcome bacterial resistance. *Org Biomiol Chem*, **9**, 4057–63.

[98] Kondo, J., Pachamuthu, K., François, B., Szychowski, J., Hanessian, S., and Westhof, E. (2007) Crystal structure of the bacterial ribosomal decoding site complexed with a synthetic doubly functionalized paromomycin derivative: a new specific binding mode to an A-minor motif enhances in vitro antibacterial activity. *ChemMedChem*, **2**, 1631–8.

[99] Baussanne, I., Bussière, A., Halder, S., Ganem-Elbaz, C., Ouberai, M., Riou, M., Paris, J.-M., Ennifar, E., Mingeot-Leclercq, M.-P., and Décout, J.-L. (2010) Synthesis and antimicrobial evaluation of amphiphilic neamine derivatives. *J Med Chem*, **53**, 119–27.

[100] Yan, R.-B., Yuan, M., Wu, Y., You, X., and Ye, X.-S. (2011) Rational design and synthesis of potent aminoglycoside antibiotics against resistant bacterial strains. *Bioorg Med Chem*, **19**, 30–40.

[101] Armstrong, E. S. and Miller, G. H. (2010) Combating evolution with intelligent design: the neoglycoside ACHN-490. *Curr Opin Microbiol*, **13**, 565–73.

[102] Pokrovskaya, V., Belakhov, V., Hainrichson, M., Yaron, S., and Baasov, T. (2009) Design, synthesis, and evaluation of novel fluoroquinolone-aminoglycoside hybrid antibiotics. *J Med Chem*, **52**, 2243–54.

[103] Hammes, G. G. (2005) *Spectroscopy for the Biological Sciences*. Wiley-Interscience.

[104] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucl Acids Res*, **28**, 235–242.

[105] Piela, L. (2007) *Ideas of Quantum Chemistry*. Elsevier Science.

[106] Atkins, P. and de Paula, J. (2005) *Physical Chemistry for the Life Sciences*. Oxford University Press.

[107] Stenta, M. and Dal Peraro, M. (2011) An introduction to quantum chemical methods applied to drug design. *Front Biosci*, **3**, 1061–78.

[108] Rassolov, V. A., Ratner, M. A., Pople, J. A., Redfern, P. C., and Curtiss, L. A. (2001) 6-31G* basis set for third-row atoms. *J Comput Chem*, **22**, 976–984.

[109] Frisch, M. J., Pople, J. A., and Binkley, J. S. (1984) Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J Chem Phys*, **80**, 3265.

[110] Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem*, **97**, 10269–10280.

[111] Jakalian, A., Bush, B. L., Jack, D. B., and Bayly, C. I. (2000) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J Comput Chem*, **21**, 132–146.

[112] Jakalian, A., Jack, D. B., and Bayly, C. I. (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem*, **23**, 1623–41.

[113] Mobley, D. L., Dumont, E., Chodera, J. D., and Dill, K. A. (2007) Comparison of charge models for fixed-charge force fields: small-molecule hydration free energies in explicit solvent. *J Phys Chem B*, **111**, 2242–54.

[114] Tsai, K.-C., Chen, Y.-C., Hsiao, N.-W., Wang, C.-L., Lin, C.-L., Lee, Y.-C., Li, M., and Wang, B. (2010) A comparison of different electrostatic potentials on prediction accuracy in CoMFA and CoMSIA studies. *Eur J Med Chem*, **45**, 1544–51.

[115] Schlick, T. (2006) *Molecular Modeling and Simulation. An Interdisciplinary Guide*. Springer.

[116] MacKerell, Jr., A. D. (2001) Atomistic Models and Force Fields. Becker, O. M., MacKerell, Jr., A. D., Roux, B., and Watanabe, M. (eds.), *Computational biochemistry and biophysics*, pp. 7–38, Taylor & Francis.

[117] Becker, O. M. (2001) Conformational Analysis. Becker, O. M., MacKerell, Jr., A. D., Roux, B., and Watanabe, M. (eds.), *Computational biochemistry and biophysics*, chap. 4, pp. 69–90, Taylor & Francis.

[118] Becker, O. M. and Watanabe, M. (2001) Dynamics methods. Becker, O. M., MacKerell, Jr., A. D., Roux, B., and Watanabe, M. (eds.), *Computational biochemistry and biophysics*, chap. 3, pp. 39–67, Taylor & Francis.

[119] Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J Chem Phys*, **98**, 10089.

[120] Rhodes, G. (2006) *Crystallography Made Crystal Clear, Third Edition: A Guide for Users of Macromolecular Models (Complementary Science)*. Academic Press.

[121] Feller, S. E., Zhang, Y., Pastor, R. W., and Brooks, B. R. (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys*, **103**, 4613.

[122] (2000) *Handbook of Chemistry and Physics, 81st Edition*. CRC Press.

[123] Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J Comput Phys*, **23**, 327–341.

[124] Amadei, A., Linssen, A. B. M., and Berendsen, H. J. C. (1993) Essential dynamics of proteins. *Proteins*, **17**, 412–25.

[125] Harte, W. E., Swaminathan, S., Mansuri, M. M., Martin, J. C., Rosenberg, I. E., and Beveridge, D. L. (1990) Domain communication in the dynamical structure of human immunodeficiency virus 1 protease. *P Natl Acad Sci USA*, **87**, 8864–8.

[126] Arunan, E., et al. (2011) Definition of the hydrogen bond (IUPAC Recommendations 2011). *Pure Appl Chem*, **83**, 1637–1641.

[127] Bloomfield, V. A., Crothers, D. M., and Tinoco, I. (2000) *Nucleic Acids: Structures, Properties, and Functions*. University Science Books.

[128] Sinnokrot, M. O., Valeev, E. F., and Sherrill, C. D. (2002) Estimates of the ab initio limit for pi-pi interactions: the benzene dimer. *J Am Chem Soc*, **124**, 10887–93.

[129] Baker, N. A., Bashford, D., and Case, D. A. (2006) Implicit Solvent Electrostatics in Biomolecular Simulation. *New Algorithms for Macromolecular Simulation*, vol. 49, pp. 263–295, Springer.

[130] Grochowski, P. and Trylska, J. (2008) Continuum molecular electrostatics, salt effects, and counterion binding–a review of the Poisson-Boltzmann theory and its modifications. *Biopolymers*, **89**, 93–113.

[131] Darden, T. A. (2001) Treatment of Long-Range Forces and Potential. Becker, O. M., MacKerell, Jr., A. D., Roux, B., and Watanabe, M. (eds.), *Computational biochemistry and biophysics*, chap. 5, pp. 91–114, Taylor & Francis.

[132] Onufriev, A., Case, D. A., and Bashford, D. (2002) Effective Born radii in the generalized Born approximation: the importance of being perfect. *J Comput Chem*, **23**, 1297–304.

[133] Dill, K. A. and Bromberg, S. (2002) *Molecular Driving Forces: Statistical Thermodynamics in Chemistry & Biology*. Garland Science.

[134] Reddy, M. R. and Erion, M. D. (2001) *Free energy calculations in rational drug design*. Kluwer Academic/Plenum Publishers, 1st edn.

[135] Schäfer, H., Mark, A. E., and van Gunsteren, W. F. (2000) Absolute entropies from molecular dynamics simulation trajectories. *J Chem Phys*, **113**, 7809.

[136] Hayward, S. (2001) Normal Mode Analysis of Biological Molecules. Becker, O. M., MacKerell, Jr., A. D., Roux, B., and Watanabe, M. (eds.), *Computational biochemistry and biophysics*, pp. 153–168, Taylor & Francis.

[137] Gohlke, H., Kiel, C., and Case, D. A. (2003) Insights into Protein–Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes. *J Mol Biol*, **330**, 891–913.

[138] Cai, Y. and Schiffer, C. A. (2010) Decomposing the energetic impact of drug resistant mutations in HIV-1 protease on binding DRV. *J Chem Theory Comput*, **6**, 1358–1368.

[139] Desheng, L., Jian, G., Yuanhua, C., Wei, C., Huai, Z., and Mingjuan, J. (2011) Molecular dynamics simulations and MM/GBSA methods to investigate binding mechanisms of aminomethylpyrimidine inhibitors with DPP-IV. *Bioorg Med Chem Lett*, **21**, 6630–5.

[140] Kaul, M. and Pilch, D. S. (2002) Thermodynamics of aminoglycoside-rRNA recognition: the binding of neomycin-class aminoglycosides to the A site of 16S rRNA. *Biochemistry*, **41**, 7695–706.

[141] Hermann, T. (2006) A-site model RNAs. *Biochimie*, **88**, 1021–6.

[142] Vicens, Q. and Westhof, E. (2001) Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site. *Structure*, **9**, 647–58.

[143] Case, D. A., et al. (2006) *AMBER 9*. University of California, San Francisco.

[144] Case, D. A., et al. (2010) *Amber 11*. University of California, San Francisco.

[145] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys*, **79**, 926.

[146] Wang, J., Cieplak, P., and Kollman, P. A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem*, **21**, 1049–1074.

[147] Hashem, Y. and Auffinger, P. (2009) A short guide for molecular dynamics simulations of RNA systems. *Methods*, **47**, 187–97.

[148] Ogle, J. M., Brodersen, D. E., Clemons, W. M., Tarry, M. J., Carter, A. P., and Ramakrishnan, V. (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science*, **292**, 897–902.

[149] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem*, **26**, 1781–802.

[150] Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994) Constant pressure molecular dynamics algorithms. *J Comput Phys*, **101**, 4177–4189.

[151] Pérez, A., Marchán, I., Svozil, D., Šponer, J., Cheatham, III, T. E., Laughton, C. A., and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J*, **92**, 3817–29.

[152] Banáš, P., Hollas, D., Zgarbová, M., Jurečka, P., Orozco, M., Cheatham, III, T. E., Šponer, J., and Otyepka, M. (2010) Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J Chem Theory Comput*, **6**, 3836–3849.

[153] Yildirim, I., Kennedy, S. D., Stern, H. A., Hart, J. M., Kierzek, R., and Turner, D. H. (2012) Revision of AMBER Torsional Parameters for RNA Improves Free Energy Predictions for Tetramer Duplexes with GC and iGiC Base Pairs. *J Chem Theory Comput*, **8**, 172–181.

[154] Vetting, M. W., Magnet, S., Nieves, E., Roderick, S. L., and Blanchard, J. S. (2004) A bacterial acetyltransferase capable of regioselective N-acetylation of antibiotics and histones. *Chem Biol*, **11**, 565–73.

[155] Fuentes-Martínez, Y., Godoy-Alcántar, C., Medrano, F., Dikiy, A., and Yatsimirsky, A. K. (2010) Protonation of kanamycin A: detailing of thermodynamics and protonation sites assignment. *Bioorg Chem*, **38**, 173–80.

[156] Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., and Baker, N. A. (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucl Acids Res*, **35**, W522–5.

[157] Li, H., Robertson, A. D., and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pKa values. *Proteins*, **61**, 704–21.

[158] Ozen, C., Malek, J. M., and Serpersu, E. H. (2006) Dissection of aminoglycoside-enzyme interactions: a calorimetric and NMR study of neomycin B binding to the aminoglycoside phosphotransferase(3′)-IIIa. *J Am Chem Soc*, **128**, 15248–54.

[159] Chen-Goodspeed, M., Vanhooke, J. L., Holden, H. M., and Raushel, F. M. (1999) Kinetic Mechanism of Kanamycin Nucleotidyltransferase from Staphylococcus aureus. *Bioorg Chem*, **27**, 395–408.

[160] Duan, Y., et al. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem*, **24**, 1999–2012.

[161] Lee, M. C. and Duan, Y. (2004) Distinguish protein decoys by using a scoring function based on a new AMBER force field, short molecular dynamics simulations, and the generalized born solvent model. *Proteins*, **55**, 620–34.

[162] Nurizzo, D., Shewry, S. C., Perlin, M. H., Brown, S. A., Dholakia, J. N., Fuchs, R. L., Deva, T., Baker, E. N., and Smith, C. A. (2003) The Crystal Structure of Aminoglycoside-3′-Phosphotransferase-IIa, an Enzyme Responsible for Antibiotic Resistance. *J Mol Biol*, **327**, 491–506.

[163] Barbieri, C. M. and Pilch, D. S. (2006) Complete thermodynamic characterization of the multiple protonation equilibria of the aminoglycoside antibiotic paromomycin: a calorimetric and natural abundance 15N NMR study. *Biophys J*, **90**, 1338–49.

[164] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field. *J Comput Chem*, **25**, 1157–74.

[165] Jakalian, A., Jack, D. B., and Bayly, C. I. (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem*, **23**, 1623–41.

[166] Frisch, M. J., et al., Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.

[167] Asensio, J. L., Hidalgo, A., Cuesta, I., González, C., Cañada, J., Vicent, C., Chiara, J. L., Cuevas, G., and Jiménez-Barbero, J. (2002) Experimental evidence for the existence of non-exo-anomeric conformations in branched oligosaccharides: NMR analysis of the structure and dynamics of aminoglycosides of the neomycin family. *Chemistry*, **8**, 5228–40.

[168] Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J Mol Graphics*, **14**, 33–38.

[169] R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[170] Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun*, **91**, 43–56.

[171] Lindahl, E., Hess, B., and van der Spoel, D. (2001) Gromacs 3.0: A package for molecular simulation and trajectory analysis. *J Mol Model*, **7**, 306–317.

[172] van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005) GROMACS: Fast, flexible, and free. *J Comput Chem*, **26**, 1701–1718.

[173] Shao, J., Tanner, S. W., Thompson, N., and Cheatham, III, T. E. (2007) Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J Chem Theory Comput*, **3**, 2312–2334.

[174] Lu, X.-J. and Olson, W. K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucl Acids Res*, **31**, 5108–21.

[175] Pérot, S., Sperandio, O., Miteva, M. A., Camproux, A.-C., and Villoutreix, B. O. (2010) Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discov Today*, **15**, 656–667.

[176] Durrant, J. D., de Oliveira, C. A. F., and McCammon, J. A. (2011) POVME: an algorithm for measuring binding-pocket volumes. *J Mol Graph Model*, **29**, 773–6.

[177] Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., and Caves, L. S. D. (2006) Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*, **22**, 2695–6.

[178] Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *P Natl Acad Sci USA*, **98**, 10037–10041.

[179] Długosz, M. and Trylska, J. (2008) Electrostatic similarity of proteins: application of three dimensional spherical harmonic decomposition. *J Chem Phys*, **129**, 015103.

[180] Srinivasan, J., Cheatham, III, T. E., Cieplak, P., Kollman, P. A., and Case, D. A. (1998) Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. *J Am Chem Soc*, **120**, 9401–9409.

[181] Lafont, V., Schaefer, M., Stote, R. H., Altschuh, D., and Dejaegere, A. (2007) Protein-protein recognition and interaction hot spots in an antigen-antibody complex: free energy decomposition identifies "efficient amino acids". *Proteins*, **67**, 418–34.

[182] Wriggers, W. and Schulten, K. (1997) Protein domain movements: detection of rigid domains and visualization of hinges in comparisons of atomic coordinates. *Proteins*, **29**, 1–14.

[183] Hinsen, K. (1998) Analysis of domain motions by approximate normal mode calculations. *Proteins*, **33**, 417–429.

[184] Hinsen, K., Thomas, A., and Field, M. J. (1999) Analysis of domain motions in large proteins. *Proteins*, **34**, 369–382.

[185] Snyder, D. A. and Montelione, G. T. (2005) Clustering algorithms for identifying core atom sets and for assessing the precision of protein structure ensembles. *Proteins*, **59**, 673–686.

[186] Kirchner, D. K. and Guntert, P. (2011) Objective identification of residue ranges for the superposition of protein structures. *BMC Bioinformatics*, **12**, 170–181.

[187] Qi, G., Lee, R., and Hayward, S. (2005) A comprehensive and non-redundant database of protein domain movements. *Bioinformatics*, **21**, 2832–2838.

[188] Poornam, G. P., Matsumoto, A., Ishida, H., and Hayward, S. (2009) A method for the analysis of domain movements in large biomolecular complexes. *Proteins*, **76**, 201–212.

[189] Aleksiev, T., Potestio, R., Pontiggia, F., Cozzini, S., and Micheletti, C. (2009) PiSQRD: a web server for decomposing proteins into quasi-rigid dynamical domains. *Bioinformatics*, **25**, 2743–2744.

[190] Potestio, R., Pontiggia, F., and Micheletti, C. (2009) Coarse-grained description of protein internal dynamics: an optimal strategy for decomposing proteins in rigid subunits. *Biophys J*, **96**, 4993–5002.

[191] Hamilton, S. W. R. and Hamilton, W. E. (1866) *Elements of quaternions (Google eBook)*. Longmans, Green, & co.

[192] Kuipers, J. B. (2002) *Quaternions and rotation sequences: a primer with applications to orbits, aerospace, and virtual reality*. Princeton University Press.

[193] Kneller, G. R. and Calligari, P. (2006) Efficient characterization of protein secondary structure in terms of screw motions. *Acta Crystallogr D*, **62**, 302–11.

[194] Keller, B., Daura, X., and van Gunsteren, W. F. (2010) Comparing geometric and kinetic cluster algorithms for molecular simulation data. *J Chem Phys*, **132**, 074110.

[195] Brooks, B. R., et al. (2009) CHARMM: The Biomolecular Simulation Program. *J Comput Chem*, **30**, 1545–1614.

[196] Campos-Olivas, R., Newman, J. L., and Summers, M. F. (2000) Solution structure and dynamics of the Rous sarcoma virus capsid protein and comparison with capsid proteins of other retroviruses. *J Mol Biol*, **296**, 633–649.

[197] Lukavsky, P. J., Kim, I., Otto, G. A., and Puglisi, J. D. (2003) Structure of HCV IRES domain II determined by NMR. *Nat Struct Biol*, **10**, 1033–1038.

[198] Ramos, A., Grünert, S., Adams, J., Micklem, D. R., Proctor, M. R., Freund, S., Bycroft, M., St Johnston, D., and Varani, G. (2000) RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J*, **19**, 997–1009.

[199] Skjærven, L., Grant, B. J., Muga, A., Teigen, K., McCammon, J. A., Reuter, N., and Martinez, A. (2011) Conformational Sampling and Nucleotide-Dependent Transitions of the GroEL Subunit Probed by Unbiased Molecular Dynamics Simulations. *PLoS Comput Biol*, **7**, e1002004.

[200] Horwich, A. L. and Fenton, W. A. (2009) Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. *Q Rev Biophys*, **42**, 83–116.

[201] Xu, Z., Horwich, A. L., and Sigler, P. B. (1997) The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature*, **388**, 741–750.

[202] Ma, J., Sigler, P. B., Xu, Z., and Karplus, M. (2000) A dynamic model for the allosteric mechanism of GroEL. *J Mol Biol*, **302**, 303–313.

[203] de Groot, B. L., Vriend, G., and Berendsen, H. J. (1999) Conformational changes in the chaperonin GroEL: new insights into the allosteric mechanism. *J Mol Biol*, **286**, 1241–1249.

[204] Kulczycka, K., Długosz, M., and Trylska, J. (2011) Molecular dynamics of ribosomal elongation factors G and Tu. *Eur Biophys J*, **40**, 289–303.

[205] Nissen, P., Kjeldgaard, M., Thirup, S., Polekhina, G., Reshetnikova, L., Clark, B. F., and Nyborg, J. (1995) Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. *Science*, **270**, 1464–1472.

[206] Abel, K., Yoder, M. D., Hilgenfeld, R., and Jurnak, F. (1996) An alpha to beta conformational switch in EF-Tu. *Structure*, **4**, 1153–1159.

[207] Polekhina, G., Thirup, S., Kjeldgaard, M., Nissen, P., Lippmann, C., and Nyborg, J. (1996) Helix unwinding in the effector region of elongation factor EF-Tu-GDP. *Structure*, **4**, 1141–1151.

[208] Réblová, K., Rázga, F., Li, W., Gao, H., Frank, J., and Šponer, J. (2010) Dynamics of the base of ribosomal A-site finger revealed by molecular dynamics simulations and Cryo-EM. *Nucleic Acids Res*, **38**, 1325–1340.

[209] Rázga, F., Zacharias, M., Réblová, K., Koča, J., and Šponer, J. (2006) RNA kink-turns as molecular elbows: hydration, cation binding, and large-scale dynamics. *Structure*, **14**, 825–835.

[210] Leontis, N. B. and Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.

[211] Romanowska, J., McCammon, J. A., and Trylska, J. (2011) Understanding the Origins of Bacterial Resistance to Aminoglycosides through Molecular Dynamics Mutational Study of the Ribosomal A-Site. *PLoS Comput Biol*, **7**, e1002099.

[212] Batey, R., Rambo, R., and Doudna, J. A. (1999) Tertiary Motifs in RNA Structure and Folding. *Angew Chem Int Edit*, **38**, 2326–2343.

[213] Horton, J. R., Ratner, G., Banavali, N. K., Huang, N., Choi, Y., Maier, M. A., Marquez, V. E., MacKerell, Jr., A. D., and Cheng, X. (2004) Caught in the act: visualization of an intermediate in the DNA base-flipping pathway induced by HhaI methyltransferase. *Nucl Acids Res*, **32**, 3877–86.

[214] Huang, N. and MacKerell, Jr., A. D. (2004) Atomistic view of base flipping in DNA. *Philos T R Soc A*, **362**, 1439–60.

[215] Sarzyńska, J., Réblová, K., Šponer, J., and Kuliński, T. (2008) Conformational transitions of flanking purines in HIV-1 RNA dimerization initiation site kissing complexes studied by CHARMM explicit solvent molecular dynamics. *Biopolymers*, **89**, 732–46.

[216] O'Connor, M., Brunelli, C. A., Firpo, M. A., Gregory, S. T., Lieberman, K. R., Lodmell, J. S., Moine, H., Van Ryk, D. I., and Dahlberg, A. E. (1995) Genetic probes of ribosomal RNA function. *Biochem Cell Biol*, **73**, 859–68.

[217] Wang, H. and Tor, Y. (1997) Electrostatic Interactions in RNA Aminoglycosides Binding. *J Am Chem Soc*, **119**, 8734–8735.

[218] Kaul, M., Barbieri, C. M., and Pilch, D. S. (2006) Aminoglycoside-induced reduction in nucleotide mobility at the ribosomal RNA A-site as a potentially key determinant of antibacterial activity. *J Am Chem Soc*, **128**, 1261–71.

[219] Asensio, J. L., Hidalgo, A., Bastida, A., Torrado, M., Corzana, F., Chiara, J. L., García-Junceda, E., Cañada, J., and Jiménez-Barbero, J. (2005) A simple structural-based approach to prevent aminoglycoside inactivation by bacterial defense proteins. Conformational restriction provides effective protection against neomycin-B nucleotidylation by ANT4. *J Am Chem Soc*, **127**, 8278–9.

[220] Matesanz, R., Diaz, J. F., Corzana, F., Santana, A. G., Bastida, A., and Asensio, J. L. (2012) Multiple keys for a single lock: the unusual structural plasticity of the nucleotidyltransferase (4′)/kanamycin complex. *Chemistry*, **18**, 2875–89.

[221] Schmitt, S., Kuhn, D., and Klebe, G. (2002) A new method to detect related function among proteins independent of sequence and fold homology. *J Mol Biol*, **323**, 387–406.

[222] Thompson, P. R., Schwartzenhauer, J., Hughes, D. W., Berghuis, A. M., and Wright, G. D. (1999) The COOH terminus of aminoglycoside phosphotransferase (3′)-IIIa is critical for antibiotic recognition and resistance. *J Biol Chem*, **274**, 30697–706.

[223] Díaz, M. D., Fernández-Alonso, M. D. C., Cuevas, G., Cañada, F. J., and Jiménez-Barbero, J. (2008) On the role of aromatic-sugar interactions in the molecular recognition of carbohydrates: A 3D view by using NMR. *Pure Appl Chem*, **80**, 1827–1835.

[224] Young, T., Abel, R., Kim, B., Berne, B. J., and Friesner, R. A. (2007) Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *P Natl Acad Sci USA*, **104**, 808–13.

[225] Michel, J., Tirado-Rives, J., and Jorgensen, W. L. (2009) Energetics of Displacing Water Molecules from Protein Binding Sites: Consequences for Ligand Optimization. *J Am Chem Soc*, **131**, 15403–15411.

[226] Shah, F., Gut, J., Legac, J., Shivakumar, D., Sherman, W., Rosenthal, P. J., and Avery, M. A. (2012) Computer-aided drug design of falcipain inhibitors: virtual screening, structure-activity relationships, hydration site thermodynamics, and reactivity analysis. *J Chem Inf Model*, **52**, 696–710.